

# A task of terawatts

The world has an abundance of renewable energy to offer, the question is how to harness it.

**T**he ultimate answer to humanity's energy problems rises every morning and sets every evening. The Sun shines down on Earth with a power of 174,000 terawatts. It lifts water from the surface of the sea to fall back down on the high mountains of the land; drives ocean currents from the tropics to the poles; spins winds around the world; and powers the green engine of life, producing practically every calorie ever consumed on Earth.

Unfortunately, humanity is now, like Hamlet, too much in the Sun. Solar energy, reduced to waste heat, no longer leaves the planet as easily as it once did. Instead it tarries in the atmosphere, thanks to the ever stronger greenhouse effect. This change in the planet's energy balance is a small one — but so great is the flow of energy from the Sun that, over decades and centuries, it will do terrible damage.

Clearly, society must slow and eventually halt the rate at which the greenhouse effect is being strengthened. The best way to achieve this, as this journal has argued on many occasions, is to increase the cost of emitting carbon dioxide and other greenhouse gases, through a cap-and-trade scheme, or possibly a direct carbon tax. But for this strategy to work there must be attractive alternative sources of energy.

The tendency among governments and traditional utilities to see renewable energy sources as oddities or add-ons is thus deeply misplaced. These sources are the future; the only question is how fast we can harness their potential. On page 816 of this issue, we survey the carbon-free technologies currently on offer. For the most part they are driven by the Sun — either directly, or via the indirect means of wind, water and plants — and it is striking to see how much they have to offer. Many of these technologies offer more than a terawatt of additional global capacity, and all have ample room for growth. Today's largest, hydropower and nuclear power, could grow by a factor of two or more — albeit with potentially troublesome non-climate consequences. But the greatest proportional capacity for growth lies with the current small-fry contributors. Wind turbines, for example, offer much greater possibilities for improvement than hydroelectric plants; the technologies for harnessing the wind far out to sea — or, for that matter, in the high currents of the jet-stream — are still comparatively immature. And even wind does not have the ultimate potential promised by the direct use of sunlight. In the future, new manufacturing methods that use alternative materials should make the capture of sunlight at least an order of magnitude cheaper than it is now.

## Ideas and incentives

The challenge is how to scale these new technologies up for a global market. A significant part of the answer is investment in focused research and development. Sadly, in most of the world, government spending on research in the energy sector has been stagnant or worse for decades. That has to change. Also crucial is an entrepreneurial culture that can get new ideas to mass markets. That requires both a social network geared to innovation and a supply of ready capital. The news

here is good: in areas such as California's Silicon Valley there is intense interest in solar energy, for example, and the industry is well positioned to exploit the same entrepreneurial vigour that has made computers and communications successful over the past few decades.

Still, enthusiastic venture capitalists will not be enough. Firms with new energy technologies need to be able to access the capital of the stock markets if they are to grow large enough to change the world. One way to make such firms attractive to investors is to guarantee them sales. This can be done through regulation — insisting on a certain amount of renewable energy in all new developments, say — through subsidy, or through direct intervention; there is nothing to stop governments from acting as bulk buyers of new capacity. Governments would be wise, though, not to indulge protectionist tendencies in the process. For investors in Texas to know that they can sell in Beijing, and for a Bangalorean to be sure of a market in Brazil, is a great benefit.

## Careful investment

Finding economically efficient ways to stimulate the growth of a particular industry is not easy. The types of subsidy that have worked to date — such as the feed-in inducements that have built Germany its solar-cell industry — are expensive even at the megawatt scale. At the scale of tens and then hundreds of gigawatts they are likely to be unsustainable. Perhaps worse than the cost, though, is the risk that poorly designed subsidies will damage markets and thwart the development they seek to encourage. However, this is a challenge to encourage technologies in a smart and flexible way, not a reason to leave the markets alone. The challenge is not how to save the world most efficiently — just how to save the world. Getting to the right place the wrong way is better than going nowhere, even if waste is involved.

There is much else to do. For example, one solution to the problem of energy storage may be found in the development of plug-in hybrid cars, which can take electricity out of a smart grid but also put it back (see page 810).

Imaginative, energetic, organized action may require new institutions. The next US president could send a strong signal of his country's seriousness by creating a well-funded Department of Renewable Energy. The development and uptake of new technologies around the world might also be encouraged by an International Renewable Energy Agency, charged with research, transmitting best practice, removing bottlenecks, and so on. But new institutions are not enough. There are interest groups that will naturally oppose this transition, and they will need to be dealt with by robust and resolute political leadership.

Even to a technoenthusiast, the scale of what needs to be done is daunting. But the wind keeps blowing, the grass keeps growing, and the Sun keeps shining. The energy is there. ■

**"The challenge is not how to save the world most efficiently — just how to save the world."**

# RESEARCH HIGHLIGHTS

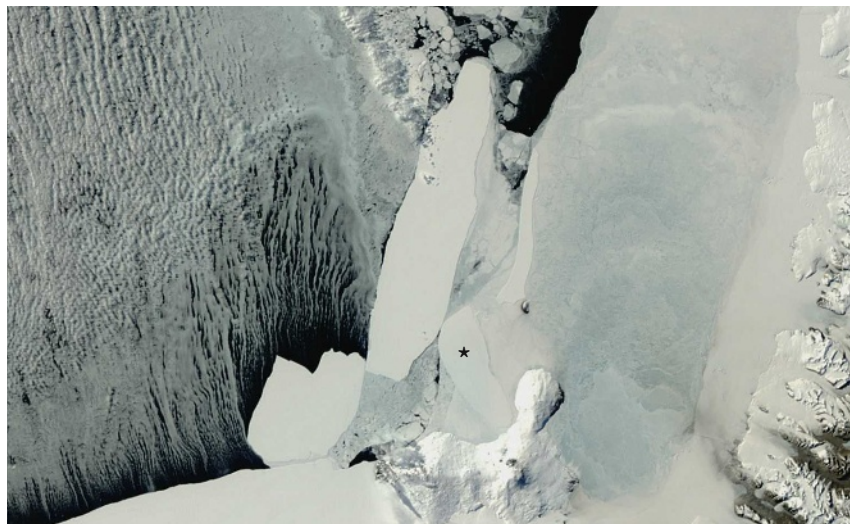
## Iceberg smash-up

*J. Geophys. Res.* doi:10.1029/2008JF001005 (2008)

When icebergs collide, they create seismic tremors that are detectable thousands of kilometres away and might one day be used to track the disintegration of ice shelves.

Douglas MacAyeal of the University of Chicago in Illinois and his colleagues placed four seismometers on a 50-kilometre-long iceberg (marked with an asterisk) that sits aground in the Ross Sea, off Antarctica. Over several weeks, they recorded many series of tremors, which they attribute to an adrift neighbouring iceberg (pictured centre) grating against the stationary one. Each series contained thousands of ice-quakes per hour.

The tremors share some similarities — but are distinct enough not to be confused — with recently discovered tremors that emanate from subduction zones and from the San Andreas fault.



D. MACAYEAL/MODIS/NASA

## COMPARATIVE BIOLOGY

### Animal models

*Proc. R. Soc. B* doi:10.1098/rspb.2008.0506 (2008)

The hierarchy of stages involved as blood stem cells develop into the various blood-cell types does not differ significantly across all species of mammal. But the rate at which these stem cells multiply does; it is faster the smaller a typical adult's mass.

This finding from David Dingli of the Mayo Clinic in Rochester, Minnesota, and his colleagues confirms that mammals used in experiments make accurate models of the human blood-cell-production process.

Dingli and his co-workers reached their conclusions after building a model of many aspects of blood-cell production across many species of mammal. They compared this with the limited experimental data available.

## PLANETARY SCIENCE

### Soft metals

*Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0804609105 (2008)

Metallic helium isn't as hard to make as scientists thought, according to Lars Stixrude of University College London and Raymond Jeanloz of the University of California, Berkeley. They calculate that the amount of squeezing needed to make the element conduct electricity is smaller when helium is hot than when it is cold; squeezing at close to 30 million atmospheres turns helium metallic at a 'mere' 20,000 kelvin.

This means that metallic helium should mix quite readily with metallic hydrogen in the fluid interiors of Jupiter and Saturn, where these conditions

are probably reached. That squelches the idea that a layer of liquid, and therefore electrically insulating, helium, floating immiscibly on metallic hydrogen, could produce helium rain in these planets' dense atmospheres.

## CLIMATOLOGY

### Winter sun

*Geophys. Res. Lett.* doi:10.1029/2008GL034160 (2008)

The weekly weather cycles detected in some big cities have been linked to higher levels of vehicle exhaust and factory emissions on weekdays than at weekends. But researchers have now found that this effect acts over a much larger area than urban centres.

Arturo Sanchez-Lorenzo from the University of Barcelona and his colleagues examined 44 years of climate data from 13 weather stations across Spain. They discovered that winter weekends tend to be drier and sunnier than weekdays. The result held for both urban and rural areas, suggesting that the hebdomadal cycles are caused by pollution's effect on regional atmospheric circulation

rather than dirty particles nucleating raindrops close to the source of smog.

The authors observed the same relationship across Western Europe, and the reverse — wetter weekends — in the eastern North Atlantic.

## ASTROPHYSICS

### Slow-motion supernovae

*Astrophys. J.* **682**, 724–736 (2008)

A survey of exploding stars shows that the farther away they are, the slower they seem to blow apart, as predicted on the basis of general relativity. The supernovae studied are known as type 1a supernovae, and are important for gauging the strength of dark energy — a mysterious force that seems to be pushing the Universe apart. These findings reassure astronomers that the far away supernovae behave like those nearby.

Stéphane Blondin at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and his colleagues studied thirteen supernovae between 3.6 and 7.5 billion light years

from Earth. The group compared several spectra from the more distant supernovae with those of nearer ones and found that the more distant explosions took longer to unfold.

## ZOOLOGY

### Predatory Lotharios

*J. Fish Biol.* **73**, 728–731 (2008)

When mothers warn their daughters that boys are after only one thing, a meal probably isn't what they have in mind. However, male fish of one species have been observed enticing females of



BLICKWINKEL/ALAMY

# RESEARCH HIGHLIGHTS

## Iceberg smash-up

*J. Geophys. Res.* doi:10.1029/2008JF001005 (2008)

When icebergs collide, they create seismic tremors that are detectable thousands of kilometres away and might one day be used to track the disintegration of ice shelves.

Douglas MacAyeal of the University of Chicago in Illinois and his colleagues placed four seismometers on a 50-kilometre-long iceberg (marked with an asterisk) that sits aground in the Ross Sea, off Antarctica. Over several weeks, they recorded many series of tremors, which they attribute to an adrift neighbouring iceberg (pictured centre) grating against the stationary one. Each series contained thousands of ice-quakes per hour.

The tremors share some similarities — but are distinct enough not to be confused — with recently discovered tremors that emanate from subduction zones and from the San Andreas fault.



D. MACAYEAL/MODIS/NASA

## COMPARATIVE BIOLOGY

### Animal models

*Proc. R. Soc. B* doi:10.1098/rspb.2008.0506 (2008)

The hierarchy of stages involved as blood stem cells develop into the various blood-cell types does not differ significantly across all species of mammal. But the rate at which these stem cells multiply does; it is faster the smaller a typical adult's mass.

This finding from David Dingli of the Mayo Clinic in Rochester, Minnesota, and his colleagues confirms that mammals used in experiments make accurate models of the human blood-cell-production process.

Dingli and his co-workers reached their conclusions after building a model of many aspects of blood-cell production across many species of mammal. They compared this with the limited experimental data available.

## PLANETARY SCIENCE

### Soft metals

*Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0804609105 (2008)

Metallic helium isn't as hard to make as scientists thought, according to Lars Stixrude of University College London and Raymond Jeanloz of the University of California, Berkeley. They calculate that the amount of squeezing needed to make the element conduct electricity is smaller when helium is hot than when it is cold; squeezing at close to 30 million atmospheres turns helium metallic at a 'mere' 20,000 kelvin.

This means that metallic helium should mix quite readily with metallic hydrogen in the fluid interiors of Jupiter and Saturn, where these conditions

are probably reached. That squelches the idea that a layer of liquid, and therefore electrically insulating, helium, floating immiscibly on metallic hydrogen, could produce helium rain in these planets' dense atmospheres.

## CLIMATOLOGY

### Winter sun

*Geophys. Res. Lett.* doi:10.1029/2008GL034160 (2008)

The weekly weather cycles detected in some big cities have been linked to higher levels of vehicle exhaust and factory emissions on weekdays than at weekends. But researchers have now found that this effect acts over a much larger area than urban centres.

Arturo Sanchez-Lorenzo from the University of Barcelona and his colleagues examined 44 years of climate data from 13 weather stations across Spain. They discovered that winter weekends tend to be drier and sunnier than weekdays. The result held for both urban and rural areas, suggesting that the hebdomadal cycles are caused by pollution's effect on regional atmospheric circulation

rather than dirty particles nucleating raindrops close to the source of smog.

The authors observed the same relationship across Western Europe, and the reverse — wetter weekends — in the eastern North Atlantic.

## ASTROPHYSICS

### Slow-motion supernovae

*Astrophys. J.* **682**, 724–736 (2008)

A survey of exploding stars shows that the farther away they are, the slower they seem to blow apart, as predicted on the basis of general relativity. The supernovae studied are known as type 1a supernovae, and are important for gauging the strength of dark energy — a mysterious force that seems to be pushing the Universe apart. These findings reassure astronomers that the far away supernovae behave like those nearby.

Stéphane Blondin at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and his colleagues studied thirteen supernovae between 3.6 and 7.5 billion light years

from Earth. The group compared several spectra from the more distant supernovae with those of nearer ones and found that the more distant explosions took longer to unfold.

## ZOOLOGY

### Predatory Lotharios

*J. Fish Biol.* **73**, 728–731 (2008)

When mothers warn their daughters that boys are after only one thing, a meal probably isn't what they have in mind. However, male fish of one species have been observed enticing females of



BLICKWINKEL/ALAMY

a closely related species to spawn so that they can consume their eggs.

Suzanne Gray of Simon Fraser University in Burnaby, Canada, and her colleagues report four instances in which a male *Telmatherina sarasinorum* was seen courting a female *T. antoniae*. The females were seen quivering, which is normally indicative of egg release. When each female finished quivering, the male immediately turned around and tried to eat her eggs.

Gray terms this behaviour 'sneaky egg-eating' and suggests it may have arisen in part because of a paucity of rich nutritional sources in Lake Matano, the Indonesian waterway where the fish live.

## HUMAN HISTORY

### Digging up data

*Proc. Natl Acad. Sci. USA* **105**, 10693–10698 (2008)  
Agriculture began in a region of the Middle East known as the Fertile Crescent.

That it diffused across Europe as human populations did is broadly accepted.

But how it moved into southern Africa has been little studied.

Brenna Henn of Stanford University, California, and her colleagues examined the frequency of two alternative versions of part of the Y-chromosome in 13 African populations. They

also compared 10 short, repeating DNA sequences known as 'microsatellites', which vary in number across populations.

The analysis shows that early farmers in what is now Tanzania bred with local populations in south-central Africa. This suggests that pastoralism and some types of cultivation were spread across the continent by people who practised them, rather than by word of mouth — and debunks the idea that farming reached southern Africa from the west.

## IMMUNOLOGY

### Successful delivery

*Cell* doi:10.1016/j.cell.2008.06.034 (2008)

Researchers have for the first time reversed symptoms of HIV infection in a living animal using the technique of RNA interference. They constructed an antibody that targets T cells — in which HIV lurks — and linked it to a peptide carrying small RNA molecules, called siRNAs. The peptide helps these siRNAs enter T cells, where they silence certain host and

virus genes crucial to the virus's replication.

Premlata Shankar, now at the Texas Tech University Health Sciences Center in El Paso, Sang-Kyung Lee of Hanyang University in Seoul, South Korea, and their colleagues injected the construct into mice genetically engineered to be easy to infect with HIV. The construct protected the mice from infection. It also restored the suppressed immune systems of mice that bore HIV-infected immune cells.

## GENETICS

### Appendages of note

*Science* **321**, 836–838 (2008)

In about 4,000 species of flowering plant, individuals can grow both male and hermaphrodite flowers. This curious trait is called andromonoecy, and it has evolved on many separate occasions. Now, Abdelhafid Bendahmane of the National Institute for Agronomic Research in Evry, France, and his

colleagues reveal the genetic mechanism behind this strategy in the melon *Cucumis melo* (pictured left).

They found that andromonoecy results from a single base change in a gene that encodes an enzyme that is involved in producing a hormone called ethylene. This hormone inhibits the development of a plant's male sex organs. Andromonoecious

melons have two copies of the altered gene, and consequently produce less enzyme. This means that any female flowers they would otherwise have grown are turned into hermaphroditic blooms.

## MATERIALS SCIENCE

### Light beat

*Soft Matter* doi:10.1039/b805434g (2008)

A polymer developed at Wright-Patterson Air Force Base in Ohio can move up and down as fast as a hummingbird's flapping wings when it is hit with laser light.

Timothy Bunning and his collaborators prepared a network of liquid crystal polymer containing azobenzene molecules, which act as linking groups. The bonds in these azo groups change reversibly from one geometrical form to another when exposed to ultraviolet laser light of a certain wavelength. This makes the polymer bend like a cantilever. As the laser's power is cranked up the cantilever moves faster. If polarizing laser light is shone on the cantilever, it moves through a different angle.



SCIENCE

## JOURNAL CLUB

Francis Albaredé

Ecole Normale Supérieure, Lyon, France

### A geochemist wonders about the Solar System's true age.

Scientists have long looked at the constituent elements of meteorites to find out how old the Sun and its planets are. The most perfect example of the oldest meteorites — those that formed at the same time as the planets — broke up and fell as a large shower near Pueblito de Allende in Mexico in 1969. This was named the Allende chondrite, and was recently the subject of a study by Jim Connelly of the University of Texas at Austin and his colleagues.

Meteorites are often dated by measuring how much aluminium-26 they contain. This isotope decays at a rate that allows researchers to tell when one meteorite is older than another, but too fast to work out these rocks' absolute ages. For the relative ages to be accurate, however, aluminium-26 must have been spread evenly among the protoplanetary debris from which meteorites were born. If it was not, this isotope would reflect where they formed as well as when they formed, and meteorite chronologies would be higgledy piggledy.

But true ages can be calculated from lead isotopes. Until recently, lead had not been measured in both of two common parts of the oldest meteorites — chondrules and calcium-aluminium-rich inclusions — for any one rock, and there was no way of telling whether different rocks formed in the same bit of the nascent Solar System. But the lead isotopes in both chondrules and calcium-aluminium-rich inclusions can be counted in Allende.

Connelly and his team have confirmed an age difference between the chondrules and calcium-aluminium-rich inclusions that had been inferred from aluminium-26 measurements. This means that the relative meteorite chronologies are correct, and that aluminium-26 was indeed evenly distributed when the Sun ignited (J. N. Connelly *et al. Astrophys. J.* **675**, L121–L124; 2008). If this is right, then the Solar System must be 4,567.5 million years old.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

a closely related species to spawn so that they can consume their eggs.

Suzanne Gray of Simon Fraser University in Burnaby, Canada, and her colleagues report four instances in which a male *Telmatherina sarasinorum* was seen courting a female *T. antoniae*. The females were seen quivering, which is normally indicative of egg release. When each female finished quivering, the male immediately turned around and tried to eat her eggs.

Gray terms this behaviour 'sneaky egg-eating' and suggests it may have arisen in part because of a paucity of rich nutritional sources in Lake Matano, the Indonesian waterway where the fish live.

## HUMAN HISTORY

### Digging up data

*Proc. Natl Acad. Sci. USA* **105**, 10693–10698 (2008)  
Agriculture began in a region of the Middle East known as the Fertile Crescent.

That it diffused across Europe as human populations did is broadly accepted.

But how it moved into southern Africa has been little studied.

Brenna Henn of Stanford University, California, and her colleagues examined the frequency of two alternative versions of part of the Y-chromosome in 13 African populations. They

also compared 10 short, repeating DNA sequences known as 'microsatellites', which vary in number across populations.

The analysis shows that early farmers in what is now Tanzania bred with local populations in south-central Africa. This suggests that pastoralism and some types of cultivation were spread across the continent by people who practised them, rather than by word of mouth — and debunks the idea that farming reached southern Africa from the west.

## IMMUNOLOGY

### Successful delivery

*Cell* doi:10.1016/j.cell.2008.06.034 (2008)

Researchers have for the first time reversed symptoms of HIV infection in a living animal using the technique of RNA interference. They constructed an antibody that targets T cells — in which HIV lurks — and linked it to a peptide carrying small RNA molecules, called siRNAs. The peptide helps these siRNAs enter T cells, where they silence certain host and

virus genes crucial to the virus's replication.

Premlata Shankar, now at the Texas Tech University Health Sciences Center in El Paso, Sang-Kyung Lee of Hanyang University in Seoul, South Korea, and their colleagues injected the construct into mice genetically engineered to be easy to infect with HIV. The construct protected the mice from infection. It also restored the suppressed immune systems of mice that bore HIV-infected immune cells.

## GENETICS

### Appendages of note

*Science* **321**, 836–838 (2008)

In about 4,000 species of flowering plant, individuals can grow both male and hermaphrodite flowers. This curious trait is called andromonoecy, and it has evolved on many separate occasions. Now, Abdelhafid Bendahmane of the National Institute for Agronomic Research in Evry, France, and his

colleagues reveal the genetic mechanism behind this strategy in the melon *Cucumis melo* (pictured left).

They found that andromonoecy results from a single base change in a gene that encodes an enzyme that is involved in producing a hormone called ethylene. This hormone inhibits the development of a plant's male sex organs. Andromonoecious

melons have two copies of the altered gene, and consequently produce less enzyme. This means that any female flowers they would otherwise have grown are turned into hermaphroditic blooms.

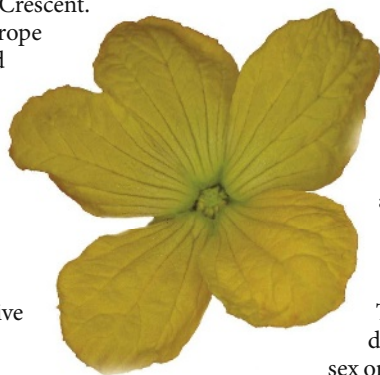
## MATERIALS SCIENCE

### Light beat

*Soft Matter* doi:10.1039/b805434g (2008)

A polymer developed at Wright-Patterson Air Force Base in Ohio can move up and down as fast as a hummingbird's flapping wings when it is hit with laser light.

Timothy Bunning and his collaborators prepared a network of liquid crystal polymer containing azobenzene molecules, which act as linking groups. The bonds in these azo groups change reversibly from one geometrical form to another when exposed to ultraviolet laser light of a certain wavelength. This makes the polymer bend like a cantilever. As the laser's power is cranked up the cantilever moves faster. If polarizing laser light is shone on the cantilever, it moves through a different angle.



SCIENCE

## JOURNAL CLUB

Francis Albaredé

Ecole Normale Supérieure, Lyon, France

### A geochemist wonders about the Solar System's true age.

Scientists have long looked at the constituent elements of meteorites to find out how old the Sun and its planets are. The most perfect example of the oldest meteorites — those that formed at the same time as the planets — broke up and fell as a large shower near Pueblito de Allende in Mexico in 1969. This was named the Allende chondrite, and was recently the subject of a study by Jim Connelly of the University of Texas at Austin and his colleagues.

Meteorites are often dated by measuring how much aluminium-26 they contain. This isotope decays at a rate that allows researchers to tell when one meteorite is older than another, but too fast to work out these rocks' absolute ages. For the relative ages to be accurate, however, aluminium-26 must have been spread evenly among the protoplanetary debris from which meteorites were born. If it was not, this isotope would reflect where they formed as well as when they formed, and meteorite chronologies would be higgledy piggledy.

But true ages can be calculated from lead isotopes. Until recently, lead had not been measured in both of two common parts of the oldest meteorites — chondrules and calcium-aluminium-rich inclusions — for any one rock, and there was no way of telling whether different rocks formed in the same bit of the nascent Solar System. But the lead isotopes in both chondrules and calcium-aluminium-rich inclusions can be counted in Allende.

Connelly and his team have confirmed an age difference between the chondrules and calcium-aluminium-rich inclusions that had been inferred from aluminium-26 measurements. This means that the relative meteorite chronologies are correct, and that aluminium-26 was indeed evenly distributed when the Sun ignited (J. N. Connelly *et al. Astrophys. J.* **675**, L121–L124; 2008). If this is right, then the Solar System must be 4,567.5 million years old.

Discuss this paper at <http://blogs.nature.com/nature/journalclub>

## NEWS

# Top climate-impacts programme shut

The lay-off last week of a senior political scientist involved in helping poor countries prepare for climate change has exposed a stark division in opinions on the core purpose of a key US climate-research institution.

The National Center for Atmospheric Research (NCAR) in Boulder, Colorado, says its hand was forced by several years of largely stagnant budgets. These have resulted in the loss of 12% of its core workforce during the past five years — ironically, during a period in which climate change moved to centre stage in Washington DC.

But the lay-off of Mickey Glantz, a high-profile researcher who has chalked up some 34 years at the institution, has raised questions about whether NCAR is turning its back on the social sciences at a time when international efforts are focusing on mitigation and adaptation. Certainly Glantz believes this is the case, saying budgets are just an excuse and that the leadership is defensively “circling the wagons”. His dismissal is tied to NCAR’s announcement last week that it is shutting its Center for Capacity Building, the highly respected outreach programme that Glantz has run since 2005.

But others associated with the programme say they believe the NCAR leadership still backs them despite its budgetary problems. “I don’t think this has anything to do with shutting down social science at NCAR,” says Linda Mearns, who recently stepped down as director of the Institute for the Study of Society



## Physicists await dark-matter confirmation

Rumours are swirling that a European satellite mission may have detected dark matter, the mysterious particles thought to make up as much of 85% of all matter in the Universe.

*Nature* has learned that the PAMELA (Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics) mission — a collaboration between Italy, Russia, Germany and Sweden — has detected a surplus of high-energy antielectrons whizzing through space. The antielectrons, also called positrons, could be the clearest signature yet of the dark matter lurking in the Milky Way, according to Dan Hooper, a theoretical physicist at Fermilab in Batavia, Illinois. “If it’s true, it’s a major discovery,” he says.

Previous space and balloon missions saw hints of the same positron surplus in the 1990s. But their energy range was limited and their measurements had high uncertainty.

PAMELA is, in principle, capable of detecting higher-energy positrons with far better accuracy than any other mission to date.

PAMELA’s principal investigator, Piergiorgio Picozza, a physicist at the University of Rome Tor Vergata, declined to comment on the findings until they have been published. However, those who have seen the PAMELA data think they are telling. “This is the type of signal that one would expect [from dark matter],” says Graciela Gelmini, a theoretical physicist at the University of California, Los Angeles, who has studied the team’s results. But Gelmini adds, the measurement is tricky, “so one has to be cautious”.

Most physicists believe that the Universe is filled with particles that rarely — if ever — interact with the stuff of stars, planets and people. This ‘dark matter’ has so far only been indirectly detected through its gravitational

pull on galaxies, and its influence on the curvature of the Universe itself.

But PAMELA may be seeing another, more concrete signature of the ghostly particles. The satellite has found an unexpected increase in the number of high-energy positrons flitting through the Galaxy, says Bob McElrath, a theorist at CERN, Europe’s high-energy physics laboratory near Geneva in Switzerland. The findings were briefly flashed across the screen at this month’s International Conference on High Energy Physics in Philadelphia, Pennsylvania, and they’ve got the community abuzz, McElrath says. “I’ve had people e-mailing and asking me to try and get the slides.”

PAMELA’s positron surplus could come from dark matter particles described by the supersymmetry theory. This predicts the existence of super-heavy counterparts to everyday



**ENVIRONMENT POLICY**  
Shake-up for Endangered  
Species Act.  
[www.nature.com/news](http://www.nature.com/news)

USFWS

## Biodiversity body 'lacks science'

Swedish researchers have launched a scathing attack on the scientific credentials of an international advisory body on biodiversity, warning that its effectiveness is being undermined by the increasing dominance of politicians and professional negotiators.

Their concerns about the work of the scientific body that advises the Convention on Biological Diversity (CBD) are widely shared, the convention's own executive secretary, Ahmed Djoghla, has told *Nature*. The convention has been signed by 168 countries who pledge to significantly reduce the current rate of biodiversity loss by 2010. Article 25 of the convention states that government representatives shall be "competent in the relevant field of expertise", but according to the Swedes, this is often not the case.

In a letter published in *Conservation Biology*, the ten scientists in the Swedish delegation to the CBD say that some parties to the convention are clearly trying to move away from science so that the convention does not interfere with trade and economic growth (L. Laikre et al. *Conserv. Biol.* 22, 814-815; 2008).

Per Wramner of Södertörn University College in Flemingsberg, who is one of the letter's authors, says that the February CBD meeting in Rome pushed them to act after it became bogged down in political wrangling and semantics. "This last meeting was a disaster from the scientific perspective," says Wramner, who chairs the Swedish government's CBD advisory group.

"Mexico and the European Union also expressed concern that there are too many new issues of procedure and of a policy nature," says Djoghla.

Conservation scientist Michael Stocking of the University of East Anglia, Norwich, UK, says that the nomination system is "the core of the problem, in that these tend to be government nominees ... not scientists who are up to date with the literature". Countries that fund the CBD will have to insist on change for it to actually happen, says Stocking, who is vice-chair of the Scientific and Technical Advisory Panel for the Global Environment Facility, which administers the funding for the CBD.

The concerns come amid attempts led by France to create a new international science policy group on biodiversity. Modelled on the same independent framework as the Intergovernmental Panel on Climate Change, this new body could mitigate some of the recently raised concerns. A 'concept note' for the new group was circulated last month by France.

Daniel Cressey

and Environment there. Mearns says social scientists within the institute will continue to work with physical scientists at NCAR on integrated research projects. "And that's the proper role for an institute in social science at NCAR."

On the other side, some scientists are questioning whether the institution has done enough to maintain, let alone build, its expertise in the physical sciences, particularly in climate modelling. These questions have been driven home by the departure of key scientists, including William Collins, who helped oversee NCAR's climate modelling programme. "I can tell you in the science divisions here, it's the worst mood people have seen in a long time, and one reflection is people walking away," says Caspar Ammann, a palaeoclimatologist at the institution.

NCAR's new director, Eric Barron, who took over in July, says the institution is in an "interesting position", caught between a dismal budgetary outlook and ongoing concerns about where NCAR should direct its limited resources. "A number of people are saying that our climate modelling programme has taken too big of a hit. People are saying very loudly that NCAR is not setting its priorities the way it should," he says. "The simple fact of the matter is that years of tight budgets are coming home to roost."

Barron says he supports the social science mission but was able to preserve several

positions throughout the institution "that are of critical importance" by eliminating a single programme that he says cost upwards of \$730,000 annually.

NCAR's base budget — almost \$88.5 million in the fiscal year 2008 — comes from the US National Science Foundation, although the institution receives significant funding from other federal agencies as well. In the fiscal year 2007, its overall budget came to \$149.3 million.

Although current appropriations bills in Congress would increase NCAR's budget, few expect this legislation to pass in an election year. Congress is likely to wind up passing a "continuing resolution" later this autumn that would effectively freeze current spending levels until at least early next year.

Roger Pielke Jr, a climate policy expert at the Center for Science and Technology Policy Research at the University of Colorado at Boulder, says it's not clear why Glantz



Mickey Glantz's social science department has been axed.

was singled out or, more broadly, how NCAR is addressing its fiscal situation. "There's really no transparency in how these decisions are made," he says.

Glantz, who has been guaranteed one year's salary, says he plans to stay on for a while, although such courtesies will not be extended to his staff, including an administrative position and two researchers.

Jeff Tollefson

particles. The lightest supersymmetrical particles are expected to be both massive and stable — making them prime candidates for dark matter. Occasionally, theorists believe, two of these particles will smash together and annihilate each other in a burst of energy. The annihilation will create a stream of more conventional particles that will eventually decay, leaving energetic electrons and positrons. In other words, the positrons detected by PAMELA could be the direct result of dark-matter annihilations.

The detection is the second dark-matter claim by Italians in recent months. In April, scientists at the DAMA/LIBRA (Dark Matter/ Large Sodium Iodide Bulk for Rare Processes) experiment, located beneath Italy's Gran Sasso mountain, claimed to have seen dark-matter particles (see *Nature* 452, 918; 2008). The PAMELA results are unconnected, and their suggested mass range for dark matter seems to contradict the DAMA claim.

PAMELA's findings are far from certain.

Identifying fast-moving positrons is extremely difficult, says Stefan Schael, a physicist at RWTH Aachen University in Germany. Their positive charge and high energies can make them look like ordinary protons, he says. Ideally, experiments would have two detectors capable of telling the difference, but PAMELA only has one. "They have only one chance for this identification," Schael says. "This is the main challenge for the group."

Even if the surplus exists, it could be from nearby astronomical sources. Objects such as neutron stars, pulsars and X-ray binary stars are capable of making energetic positrons that would fool PAMELA. "When the data are released there will certainly be a large number of people looking at the question," says Hooper.

For now, however, the community is waiting with bated breath. As McElrath says: "We all wonder what's going on up there."

Geoff Brumfield



**ENVIRONMENT POLICY**  
Shake-up for Endangered  
Species Act.  
[www.nature.com/news](http://www.nature.com/news)

USFWS

## Biodiversity body 'lacks science'

Swedish researchers have launched a scathing attack on the scientific credentials of an international advisory body on biodiversity, warning that its effectiveness is being undermined by the increasing dominance of politicians and professional negotiators.

Their concerns about the work of the scientific body that advises the Convention on Biological Diversity (CBD) are widely shared, the convention's own executive secretary, Ahmed Djoghla, has told *Nature*. The convention has been signed by 168 countries who pledge to significantly reduce the current rate of biodiversity loss by 2010. Article 25 of the convention states that government representatives shall be "competent in the relevant field of expertise", but according to the Swedes, this is often not the case.

In a letter published in *Conservation Biology*, the ten scientists in the Swedish delegation to the CBD say that some parties to the convention are clearly trying to move away from science so that the convention does not interfere with trade and economic growth (L. Laikre et al. *Conserv. Biol.* 22, 814-815; 2008).

Per Wramner of Södertörn University College in Flemingsberg, who is one of the letter's authors, says that the February CBD meeting in Rome pushed them to act after it became bogged down in political wrangling and semantics. "This last meeting was a disaster from the scientific perspective," says Wramner, who chairs the Swedish government's CBD advisory group.

"Mexico and the European Union also expressed concern that there are too many new issues of procedure and of a policy nature," says Djoghla.

Conservation scientist Michael Stocking of the University of East Anglia, Norwich, UK, says that the nomination system is "the core of the problem, in that these tend to be government nominees ... not scientists who are up to date with the literature". Countries that fund the CBD will have to insist on change for it to actually happen, says Stocking, who is vice-chair of the Scientific and Technical Advisory Panel for the Global Environment Facility, which administers the funding for the CBD.

The concerns come amid attempts led by France to create a new international science policy group on biodiversity. Modelled on the same independent framework as the Intergovernmental Panel on Climate Change, this new body could mitigate some of the recently raised concerns. A 'concept note' for the new group was circulated last month by France.

Daniel Cressey

and Environment there. Mearns says social scientists within the institute will continue to work with physical scientists at NCAR on integrated research projects. "And that's the proper role for an institute in social science at NCAR."

On the other side, some scientists are questioning whether the institution has done enough to maintain, let alone build, its expertise in the physical sciences, particularly in climate modelling. These questions have been driven home by the departure of key scientists, including William Collins, who helped oversee NCAR's climate modelling programme. "I can tell you in the science divisions here, it's the worst mood people have seen in a long time, and one reflection is people walking away," says Caspar Ammann, a palaeoclimatologist at the institution.

NCAR's new director, Eric Barron, who took over in July, says the institution is in an "interesting position", caught between a dismal budgetary outlook and ongoing concerns about where NCAR should direct its limited resources. "A number of people are saying that our climate modelling programme has taken too big of a hit. People are saying very loudly that NCAR is not setting its priorities the way it should," he says. "The simple fact of the matter is that years of tight budgets are coming home to roost."

Barron says he supports the social science mission but was able to preserve several

positions throughout the institution "that are of critical importance" by eliminating a single programme that he says cost upwards of \$730,000 annually.

NCAR's base budget — almost \$88.5 million in the fiscal year 2008 — comes from the US National Science Foundation, although the institution receives significant funding from other federal agencies as well. In the fiscal year 2007, its overall budget came to \$149.3 million.



**Mickey Glantz's social science department has been axed.**

Although current appropriations bills in Congress would increase NCAR's budget, few expect this legislation to pass in an election year. Congress is likely to wind up passing a "continuing resolution" later this autumn that would effectively freeze current spending levels until at least early next year.

Roger Pielke Jr, a climate policy expert at the Center for Science and Technology Policy Research at the University of Colorado at Boulder, says it's not clear why Glantz

was singled out or, more broadly, how NCAR is addressing its fiscal situation. "There's really no transparency in how these decisions are made," he says.

Glantz, who has been guaranteed one year's salary, says he plans to stay on for a while, although such courtesies will not be extended to his staff, including an administrative position and two researchers.

Jeff Tollefson

particles. The lightest supersymmetrical particles are expected to be both massive and stable — making them prime candidates for dark matter. Occasionally, theorists believe, two of these particles will smash together and annihilate each other in a burst of energy. The annihilation will create a stream of more conventional particles that will eventually decay, leaving energetic electrons and positrons. In other words, the positrons detected by PAMELA could be the direct result of dark-matter annihilations.

The detection is the second dark-matter claim by Italians in recent months. In April, scientists at the DAMA/LIBRA (Dark Matter/ Large Sodium Iodide Bulk for Rare Processes) experiment, located beneath Italy's Gran Sasso mountain, claimed to have seen dark-matter particles (see *Nature* 452, 918; 2008). The PAMELA results are unconnected, and their suggested mass range for dark matter seems to contradict the DAMA claim.

PAMELA's findings are far from certain.

Identifying fast-moving positrons is extremely difficult, says Stefan Schael, a physicist at RWTH Aachen University in Germany. Their positive charge and high energies can make them look like ordinary protons, he says. Ideally, experiments would have two detectors capable of telling the difference, but PAMELA only has one. "They have only one chance for this identification," Schael says. "This is the main challenge for the group."

Even if the surplus exists, it could be from nearby astronomical sources. Objects such as neutron stars, pulsars and X-ray binary stars are capable of making energetic positrons that would fool PAMELA. "When the data are released there will certainly be a large number of people looking at the question," says Hooper.

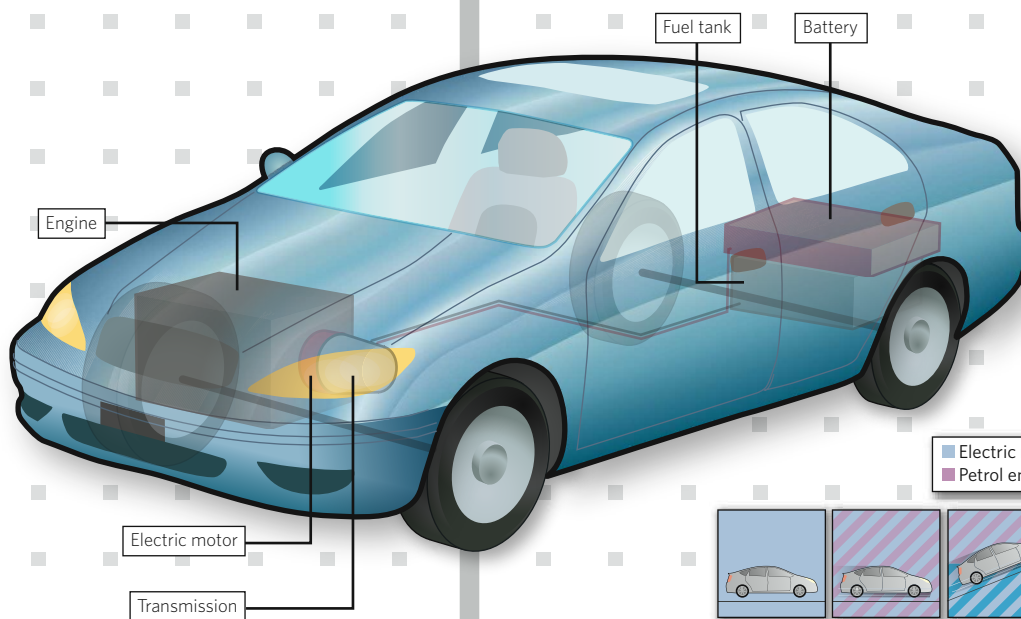
For now, however, the community is waiting with bated breath. As McElrath says: "We all wonder what's going on up there."

Geoff Brumfield

# FOUR WHEELS GOOD?

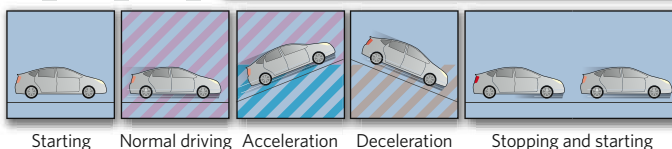
With the world's love of cars showing little sign of abating, manufacturers are under increasing pressure to make vehicles less polluting and oil dependent. **Duncan Graham-Rowe** explores some of the technologies that could keep us on the road.

## HYBRIDS

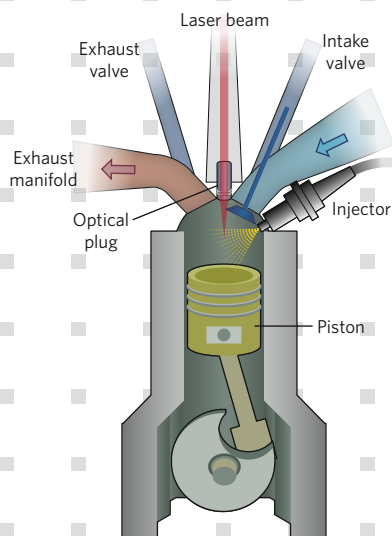


One way to reduce the energy cars consume is to recycle it. Hybrid vehicles use the combustion engine to charge up a battery that provides extra power to the engine when accelerating, thus reducing fuel consumption. 'Regenerative braking', in which the electric motor provides resistance to the drive train to help slow the vehicle, converts the kinetic energy into electricity, which is then stored for later use. The benefits of hybrids can really be seen during town driving — the fuel savings may be only a few per cent on longer trips. Fleets of hybrids could also provide a place to store electricity on a large scale, helping in the use of various renewable technologies (see page 805).

■ Electric motor in use throughout ■ Battery power used  
■ Petrol engine used ■ Battery recharging



## LASER INJECTION



Replacing spark plugs with laser-pulse ignition systems may sound over the top, but it makes sense, says Andrew Scarisbrick, supervisor for UK government and university collaborations at Ford Motor Company's Dunton Research and Engineering Centre near Basildon. Laser ignition systems can reduce fuel consumption and emissions because they give better control of where in the cylinder the spark occurs. Ideally, the spark should be as far from the cylinder walls as possible, so that the flame front is less able to form harmful nitrous oxides. But this technology is still a good few years off, Scarisbrick says. "At the moment the cost is astronomic."

## ENGENIUS

One goal is to make existing combustion engines more efficient. Last year, the gauntlet was thrown down with the X-Prize Foundation's Automotive X Prize, a US\$10-million bounty for whoever can develop a car that can do 100 miles per gallon.

Automotive engineering company Ricardo UK, based in Sussex, has made some headway, designing an engine that can switch between two-stroke and four-stroke modes. The company claims this could reduce fuel consumption by 27% and emissions by a similar amount. Four-stroke engines carry out the four stages of air intake, compression, combustion and exhaust in four strokes of a piston, whereas two strokes take just two strokes of the piston. Four strokes are more fuel-efficient at constant high speeds, whereas two-strokes are more efficient when accelerating or pulling a heavy load. Software in Ricardo's concept engine, 2/4 SIGHT, controls hydroelectric valves that switch between four-stroke mode during cruising to two-stroke mode when accelerating or hill climbing.

## BETTER BATTERIES

One reason electric vehicles have never taken off is the lack of decent battery technologies. Lithium-ion batteries are currently the best candidates for cars because they deliver power much more quickly than others. But they only carry enough charge to last for around 100 miles before needing recharging. They also have to be replaced every couple of years. And in extreme circumstances they can catch fire or explode. So, many car

manufacturers, including Toyota with the Prius hybrid vehicle, have opted for nickel metal hydride batteries, which pack less punch than the lithium-ion ones but last longer and are safer.

Companies such as Massachusetts-based A123Systems are finding new ways to squeeze more life out of a battery. The cathodes within batteries degrade with repeated use and recharging. So A123Systems has developed a birdcage-like

nanosstructure of lithium iron phosphate surrounding the cathode that prevents it from expanding and contracting during charging and discharging, and so stops the cathode wearing down so quickly. According to the company, its batteries have a potential lifetime ten times that of conventional ones. And although they are not as powerful as the regular lithium-ion batteries, they are a vast improvement on the nickel metal hydride type.

## CONTROL FREAK

Even subtle alterations to the way cars handle on roads can improve fuel efficiency. Drive-by-wire technology would remove the mechanical linkages between the controls and the brakes, throttle and steering mechanisms. Computers continually interrogate sensors about the car's handling and the road conditions to calculate, for example, the optimal throttle position at that moment. Most countries currently require a mechanical linkage between the steering wheel and wheels themselves, so manufacturers would have to persuade governments that the technology is safe.

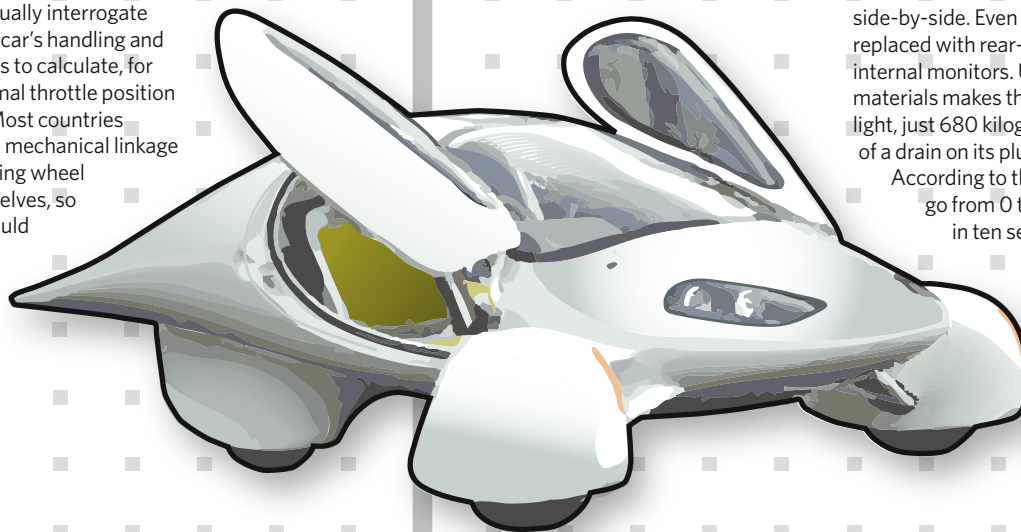
## BETTER BY DESIGN

The electric two-seater renewable energy vehicle (TREV), developed at the University of South Australia in Adelaide, was designed to carry only two people because, according to its creators, 90% of urban trips

require no more than two seats.

Thinking along similar lines, the Californian company Aptera recently launched a two-seater electric car (pictured) that it claims has the most energy-efficient and lowest-drag shape that can surround two people sitting side-by-side. Even its wing mirrors are replaced with rear-facing cameras and internal monitors. Use of composite materials makes the three-wheeler very light, just 680 kilograms, placing less of a drain on its plug-in electric motor.

According to the company, it can go from 0 to 60 miles per hour in ten seconds and has a top speed of 85 mph.



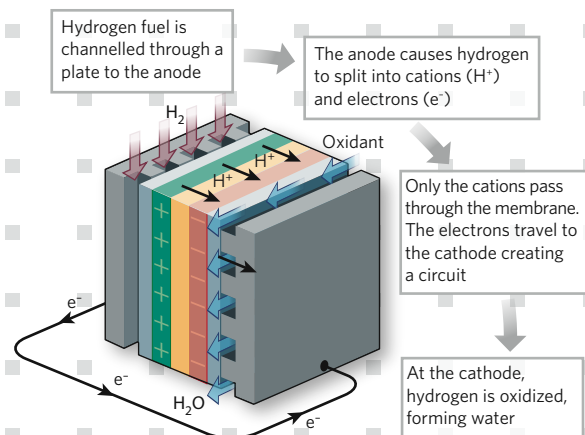
## HYDROGEN FUEL

Widely tipped to replace oil as a transportation fuel, hydrogen has nearly three times the energy density of petrol by mass, and when used to power fuel cells it produces only steam as a waste product. Most large car manufacturers are developing some form of hydrogen-powered car.

But there are major obstacles to overcome before hydrogen could become a mainstream fuel — how to produce, transport and store it. "Hydrogen requires a huge infrastructure change that's not likely to happen for a long time," says David Sims-Williams, an engineer at Durham University, UK. Producing it using renewable sources of energy is extremely inefficient, and the vast majority of hydrogen is currently produced from coal or natural gas, which is less than ideal.

Another major problem is that, even in liquid form, hydrogen has a tenth the density of water. So hydrogen has less than a third of the energy density of petrol by volume. As things stand, a car's fuel tank would not only have to be cooled to below  $-250^{\circ}\text{C}$  to keep the hydrogen liquid but would also have to be many times larger than existing ones.

In 2003, the US Department of Energy issued a challenge to scientists to develop new materials that can store enough hydrogen to make up 6% of the material's total mass. Frantisek Svec at Lawrence Berkeley National Laboratory in California has developed a nanoporous polymer that adsorbs hydrogen atoms reversibly on its surface, allowing much higher densities to be stored. But the hydrogen is only 1.5% of the polymer's mass, still way short of the target.



# The AIDS fight: looking ahead to 2010

As this month's International AIDS Conference in Mexico City ends, **Erika Check Hayden** asks leaders of the HIV community what they expect to see on the agenda when the conference reconvenes in Vienna in 2010 — the year by which the world has agreed to provide universal access to HIV treatment.

## Julio Montaner

*President, International AIDS Society, 2008–10*

We want to make 2010 an accountability conference, so that we have a clear idea of which promises have been made, kept and broken in the roll-out of antiretroviral therapy. And there will be more focus on eastern Europe, where the epidemic is a catastrophe brewing. The issue of intravenous drug users is going to become dominant. Also, the distance between what we know works and what we're actually doing needs to be closed, or the epidemic will become totally out of control.



## Tachi Yamada

*President, Bill & Melinda Gates Foundation Global Health Program*

I hope we will have thought through how to deal with drug resistance and the need to go to second-line treatments. In prevention, I would like to see that we've started on the roll-out of circumcision, and have doubled the prevention of mother-to-child transmission. I'd like to know whether the addition of antiretrovirals to microbicide treatment will make a difference. And I'd like a plan to address multiple concurrent partners in high-prevalence countries.



## Mitchell Warren

*Executive director, AIDS Vaccine Advocacy Coalition*

We should have data from up to four of the initial PrEP studies and from a 16,000-person Thai vaccine trial of a prime-boost combination of ALVAC-HIV and AIDSVAX. Ideally, we will have at least one new vaccine and/or microbicide product in an efficacy trial. We should be talking about impact from the male-circumcision programme scale-up as opposed to research results. And we will be talking about 'antiretroviral-based prophylaxis' as opposed to microbicides or PrEP.



## Michel Kazatchkine

*Executive director, Global Fund to Fight AIDS, Tuberculosis and Malaria*

I expect more progress from the science of treatment, with more powerful combination treatments and ways of selecting combinations of antiretrovirals to provide long-term suppression of the virus and avoid resistance. I expect a lot more information on how antiretroviral therapy can help with prevention. And I'm looking for what percentage of countries will have reached universal coverage in treatment; I'm quite hopeful for a number of middle-income countries in western Africa and Latin America.



## Zeda Rosenberg

*Chief executive, International Partnership for Microbicides*

The two ongoing trials of early-generation microbicides will have been completed and the next generation of microbicides should be well on its way to showing efficacy. The CAPRISA trial of tenofovir gel should be finished, and the data will be under analysis. Several trials of next-generation microbicides are due to be launched — both safety studies of newer products, and efficacy studies of products based on non-nucleoside reverse transcriptase inhibitors.



## Mark Dybul

*US global AIDS coordinator*

I hope we have more data on the impact of HIV on countries' health systems and on innovative approaches to incorporating the private sector into prevention. The HIV community should be looking at other public-health efforts that have had generational behavioural change. We should see more full-country ownership, with all sectors talking about what they've done with the freedom to do what they need to do. I hope we see more of that and less discussion of global architecture and the things we know how to do.



## Peter Piot

*Executive director, Joint United Nations Programme on HIV/AIDS*

I think even more attention will be paid to the 'how' of prevention. We should have results of the big studies of pre-exposure prophylaxis [PrEP]. I hope we will have a better understanding of protective immunity and more information on resistance to antiretrovirals. And there will be more on linking up countries' health systems to AIDS activities.



## Alan Bernstein

*Executive director, Global HIV Vaccine Enterprise*

The revolution in biomedical research, with high-throughput technology, systems biology and computational biology, needs to be applied to vaccines. From that we might learn how to achieve protection. There needs to be more emphasis on how we measure what's going on early in infection. There seems to be a brief window of opportunity after someone becomes infected when we can intervene.



## Kevin De Cock

*Director of the HIV/AIDS Department, World Health Organization*

Male circumcision needs to be scaled up. We're waiting for answers about the use of antiretroviral therapy in people already infected with HIV, to see whether it lowers viral load and how that translates into reduced transmission. I don't think we're going to get universal access in all countries by 2010, but there are some places, such as Namibia and Botswana, that can do it.



FROM LEFT: R. JENSEN/AP; INT. PARTNERSHIP MICROBICIDES; J.S. APPLEWHITE/AP

FROM LEFT: INT. AIDS SOC.; BILL & MELINDA GATES FOUND.; UNIV. WISCONSIN-MADISON

FROM LEFT: C. ARANDA/MONDAPHOTO/INT. AIDS SOC.; L. TREZZINI, KEYSTONE/AP



**FRAUD AND MISCONDUCT**  
Do we need a penal system for science?  
[www.nature.com/news](http://www.nature.com/news)

PUNCHSTOCK

# Anthrax case ignites new forensics field

The messy tragedy surrounding the 2001 US anthrax attacks and the suicide of chief suspect Bruce Ivins has thrown the emergent field of microbial forensics into the spotlight. The forensic techniques proved vital in allowing the Federal Bureau of Investigation (FBI) to make its case that the anthrax used in the attacks came from a particular sample in Ivins's lab.

Ivins died before facing charges, but had his case reached trial it would have been a major test for the discipline, which barely existed before 2001.

Microbial forensics traces the origin of a biological agent using a range of biochemical analyses, including genomic sequencing and protein and carbohydrate fingerprinting. The discipline has its roots in military and intelligence communities, and was used successfully in court as far back as 1998, in the trial of a doctor in Louisiana who intentionally infected his former mistress with HIV. But the field took off during the anthrax investigation with an



**Anthrax spores were forensically analysed.**

influx of government funding and the advent of cheaper, faster sequencing technology.

Microbial forensics is likely to have an increasing role beyond biocrime. "The big thing that's going to be useful out of all this is its application to molecular epidemiology," says Paul Keim, a molecular geneticist at Northern Arizona University in Flagstaff, who worked with the FBI investigation. For example, microbial forensics could be used to pinpoint the source of food- or

water-borne pathogens. Referring to the recent US *Salmonella* outbreak, Keim says, "If the same thing happens in a year, I would expect to see hundreds of *Salmonella* genomes sequenced."

The field could also be important in lawsuits over hospital-acquired infections, says microbiologist Abigail Salyers of the University of Illinois at Urbana-Champaign. People have sued hospitals for millions, claiming they contracted methicillin-resistant *Staphylococcus aureus* (MRSA) as patients. Microbial forensics could be used to trace an infection to its source and ascertain whether it was acquired at a particular hospital.

Microbiologists will need to be confident of their technique — forensic analysis requires unusual rigour. In a journal publication, it is usually understood that further work is needed, says Keim. "In court, you don't have that luxury," he says. "You have to get it right the first time."

**Amber Dance**

J. HANEY CARR/CDC

## Large Hadron Collider passes first proton test

The world's largest particle accelerator has tasted its first protons.

On 8 August, physicists injected a few billion protons into a section of the Large Hadron Collider (LHC) at CERN, Europe's high-energy physics laboratory near Geneva, Switzerland. The accelerator will eventually drive trillions of protons into each other at energies high enough to perhaps generate new kinds of particles.

The small beam tested the synchronization between the LHC and a booster accelerator. "I must say I was very pleased," says Lyn Evans, LHC project leader at CERN. "It went extremely smoothly."

Further tests will be carried out in coming weeks, with a view to having protons running laps around the 27-kilometre LHC ring by 10 September.

## Environment agency denies plea to cut biofuel goals

The US Environmental Protection Agency (EPA) last week upheld the national mandate to use biofuels. The decision was made despite opposition from livestock and environmental groups, which say that the expansion of corn (maize) ethanol is pushing up food prices and damaging the environment.

EPA administrator Stephen Johnson denied the request of Rick Perry, the Republican governor of Texas, to halve the 'renewable fuels standard', which requires the nation to produce 34 billion litres of ethanol in 2008. Perry had argued that livestock producers in his state spent an additional US\$1.2 billion on corn feed between 2004 and 2007, a trend that he

linked to the expansion of biofuels.

But Johnson says that repealing the standard would do little, if anything, to bring corn prices down. He also says that the EPA's analysis found no evidence that the mandate is damaging the national economy.

For a longer version of this story, see <http://tinyurl.com/5hr3jv>.

## San Francisco's biotech hub gets lift from Pfizer deal

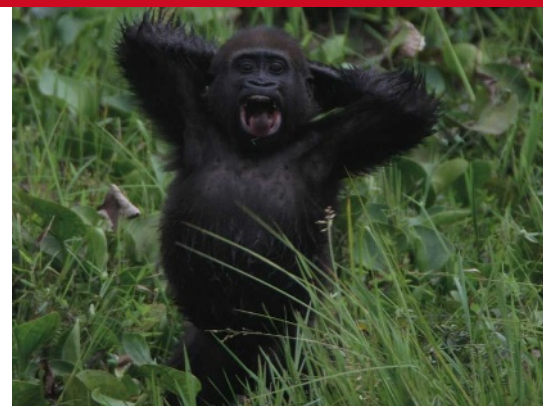
San Francisco's drive to turn the Mission Bay area into a technology-transfer hub has been boosted by Pfizer's decision to move one of its research centres to the site.

In early 2010, the Biotherapeutics and Bioinnovation Center will move into newly constructed lab facilities near the Mission Bay research complex run by the University of California, San Francisco. About 100 of the centre's employees will move there from its current base in South San Francisco.

Five years after the university opened facilities at Mission Bay, biotech firms are finally trickling in. In June, Pfizer, which is headquartered in New York, agreed a US\$9.5-million, three-year deal for research collaborations with Mission Bay's California Institute for Quantitative Biosciences.

## Gorilla numbers doubled in the Congo

Western lowland gorillas (*Gorilla gorilla gorilla*) are more common than previously thought, according to a census of the northern regions of the Republic of the Congo. Led by the New-York-based Wildlife Conservation Society (WCS), the count found 125,000 of the apes in a 47,000-square-kilometre area. Previous estimates held that there were fewer than 100,000



western lowland gorillas in the entire world, and the International Union for Conservation of Nature has speculated that the Ebola virus could have reduced numbers even further.

Census-takers counted the gorillas' sleeping nests. The WCS attributes the high numbers to successful management of protected areas in the Republic of the Congo, a food-rich habitat and the "remoteness and inaccessibility" of the region. The work was presented at the International Primatological Society Congress in Edinburgh, UK.

Even so, a separate global assessment presented at the same conference reports that half of primate species are threatened with extinction in the next ten years.

For a longer version of this story, see <http://tinyurl.com/68ewes>

## FDA further tightens conflict of interest policies

The US Food and Drug Administration (FDA) is introducing guidelines to reduce potential conflicts of interest among members of its advisory committees.

In part, the regulations address a 2007 law that requires the agency to cut the number of advisers who have conflicts of interest by one-quarter over the next five years. According to the guidelines, committee members may not have financial ties to the topics under consideration. Those with financial interests totalling less than \$50,000 would be eligible for waivers allowing them to participate.

The guidelines also instruct members of committees to vote simultaneously. Previously, members stated their votes aloud in succession, which could influence later voters. In addition, the FDA intends to make information about committee activities more accessible by posting related documents on its website prior to meetings.

### SCIFOO BLOGGED

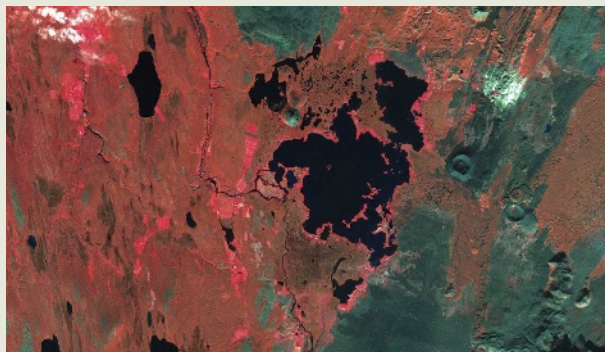
Last week saw the 2008 SciFoo, the annual 'unconference' organized by *Nature*, Tim O'Reilly and Google. Topics ranged from the Large Hadron Collider to the neuroscience of the brain-machine interface and a new social contract for science. Details and participants are at [www.nature.com/scifoo](http://www.nature.com/scifoo).

## Hotspot for midges proves to be fertile ground

This false-colour infrared image, taken in September 2006 with the ASTER sensor on NASA's Terra satellite, shows Mývatn, a shallow lake in Iceland from which emerge huge numbers of mating midges in summer.

Bright red indicates vigorous vegetation; the thin border of hot pink around the lake has been fertilized by the bodies of dead midges. In late July and early August, 135 kilograms of biomass was deposited there per hectare per day, according to Phil Townsend, a remote sensing specialist at the Department of Forest and Wildlife Ecology at the University of Wisconsin in Madison.

Townsend, who presented his results last week at the Ecological Society of America meeting in Milwaukee, Wisconsin, is using satellite imagery to identify more 'midge lakes'.



P. TOWNSEND/ASTER

T. BREUER/WILDLIFE CONSERVATION SOCIETY/MP/EVOL. ANTHROPOL.

## BOOKS &amp; ARTS



SOLEIL NOIR/PHOTONONSTOP/PHOTOLIBRARY

## On the scent

For Marcel Proust, the madeleine evoked childhood memories — a new treatment of the science of smell attempts to take our everyday experience of odour to a more insightful level, explains **Gary Beauchamp**.

### What the Nose Knows: The Science of Scent in Everyday Life

by Avery Gilbert

Crown Publishers: 2008. 304 pp.

£15.99, \$23.95

In 2004, the Nobel Prize for Physiology or Medicine was awarded to Linda Buck and Richard Axel for their work on the science of smell. They discovered the family of more than a thousand olfactory receptors that recognize the volatile compounds that make up our scented world. To researchers in the field, this honour was neither surprising nor unanticipated. Many outsiders wondered: what is the social, biological and medical significance of this breakthrough that makes it so noteworthy?

Avery Gilbert's engaging book *What the Nose Knows* attempts to answer this question. Focusing on our experiences of scent in everyday life, it proclaims the importance and curious nature of the human sense of smell. Until the last chapter, which speculates on the future, Gilbert has little to say about the biological science underlying olfaction. He has even less to say about odours in other mammals and insects, where they have a greater role in regulating social and

sexual activity than in humans. What this book lacks in basic biological science, it makes up for in psychological analysis, complemented throughout with literary allusions that Gilbert skilfully uses to illustrate his insights.

Gilbert draws on his personal experience, both anecdotal and scientific. He asks how many odours there are and how we might determine them, and describes the principles of perfumery. He addresses the psychological aspects of odour phobias and multiple chemosensory disorder, and explores odour memory within the literature of Marcel Proust. Gilbert seems especially fond of stories of the smell of dead bodies and human gaseous emanations. The book is perhaps overly larded with short, albeit fascinating, stories and olfactory anecdotes, and lighter on synthesis. Citations to the psychological literature abound, but they are listed at the end and the reader must flick back and forth to find a source.

The story of odorizing the movie experience is delightfully told. The 'battle of the smellies' between the scientifically based Smell-O-Vision, pushed by the impresario Mike Todd, and its rival AromaRama illuminates both the entrepreneurial spirit of the 1950s and the scientific and technological aspects of odour choice and

delivery. The difficult task of odour removal is highlighted by cinema that, in the words of one reviewer, was filled with scent "strong enough to give a bloodhound a headache".

Gilbert adopts a strong position on matters of controversy, sometimes trying hard to arouse disagreement. He persuasively dismantles the idea of Proust as poster boy for the profundity of odour memory, but I remain convinced that major research efforts to understand the phenomenon are still needed. When I encounter the scent of my grandfather's garage, I am not only reminded of the place — I am briefly transported there. Gilbert, too, calls attention to this power: "Like a nightclub mentalist, the mind presents us with a memory it picked from our pocket when we weren't looking." The neural pathways unique to olfaction are probably involved but do not completely explain the experience.

Similarly, I question Gilbert's argument that smell makes a greater contribution to flavour than does taste. First, it is unclear what metric would be useful in measuring this. And there are many reasons to believe that taste — sweet, sour, salty, bitter and umami or savoury — is more potent as far as food selection and intake are concerned. Smell permits one to discriminate strawberry from cherry, but taste provides

the information necessary to decide whether an object should be taken into the body or expelled.

Many years ago, when working with Gilbert at the Monell Chemical Senses Center in Philadelphia, I recognized his way with words as

well as with scientific research. *What the Nose Knows* melds the academic and business worlds of smell into an entertaining and illuminating rumination on this almost magical sense that, even with a Nobel Prize to its credit, still holds many mysteries. ■

**Gary Beauchamp** is director of the Monell Chemical Senses Center, 3500 Market Street, Philadelphia 19104, Pennsylvania, USA. With Stuart Firestein and David V. Smith, he is co-editor of *Olfaction and Taste*.  
e-mail: beauchamp@monell.org

## Playing to win

**Starting With Serotonin: How a High-Rolling Father of Drug Discovery Repeatedly Beat the Odds**

by Ann G. Sjoerdsma

Improbable Books: 2008. 640 pp. \$27.50

Science is a gamble. Publication, applying for grants, student admissions and corporate relationships all involve high-stakes bets, a mixture of skill and luck, and often a bit of bluffing. Which game is science most like? It's not a slot machine, mindlessly addictive. In dark moments it may seem like roulette, with its powerful house advantage and long odds. Sometimes it's a horse race, when one thoroughbred laboratory noses out another in isolating a long-sought gene or subatomic particle. For Albert Sjoerdsma, sometimes called the father of clinical pharmacology, science was most like craps.

Craps, an intricate dice game that can involve many players and interweaving rounds of betting, is a thinking person's pastime. Winning depends on an understanding of probability and being able to weigh complex constellations of risks and payouts. In *Starting with Serotonin*, Sjoerdsma's biographer daughter Ann Sjoerdsma argues that craps was her father's favourite game of chance — and the key theme in his scientific life.

Albert Sjoerdsma came to the table with a modest stack of chips. Born in 1924 and raised near Chicago in Illinois, he grew up with little money or social sophistication, but with a first-class brain and a mountain of confidence. He was a rough-and-tumble child, more likely to be found playing sports or getting into mischief than curled up with his nose in a book. Yet his grades were nearly perfect, and he was accepted at the University of Chicago under president Robert Hutchins, whose innovative programmes helped to train some of the best minds of the late twentieth century. Sjoerdsma was cocky and sometimes disrespectful, especially when faced with arbitrary displays of power. His daughter describes him as a maverick, a clichéd but apt term.

Rather than fold with a bachelor's degree, Sjoerdsma stayed in the game, taking an MD and a PhD at Chicago. He then went east to Bethesda, Maryland, where he joined the National Institutes of Health (NIH) in 1951. After two years of residency at the Public Health Service's Marine Hospital in Baltimore, Sjoerdsma landed a position at the National Heart Institute back in Bethesda. There, he formed a team that became known locally as the "wild bunch", a group of brilliant, hard-working and hard-playing researchers.

Sjoerdsma began exploring ways of reducing high blood pressure, leading to his investigation of the then recently discovered neurotransmitter serotonin. His analysis of its effects on different organ systems and metabolic pathways led him into a strongly applied style of pharmacology, in which he largely eschewed lab-based studies in favour of whole patients, and focused on bridging the gap between laboratory science and clinical medicine.

He was a pioneer in the development of monoamine oxidase inhibitors as antidepressants, such as iproniazid, originally developed as an anti-tuberculosis drug. With colleague Sidney Udenfriend, he found that monoamine oxidase was a major pathway for serotonin in both mice and humans, and that inhibitors such as iproniazid raised blood serotonin levels. This established the physiological basis of the antidepressant action of these drugs. Sjoerdsma's

20 years at NIH coincided with the 'golden age' of its intramural research, an era of Nobel prizes, headline-grabbing breakthroughs and major contributions to science.

In 1972, he parlayed his successes as a bench scientist into a job as director of a new research centre in Strasbourg, France, set up by the pharmaceutical company Richardson-Merrell. His blunt, incisive intellectual and administrative style was polarizing in genteel Europe. He

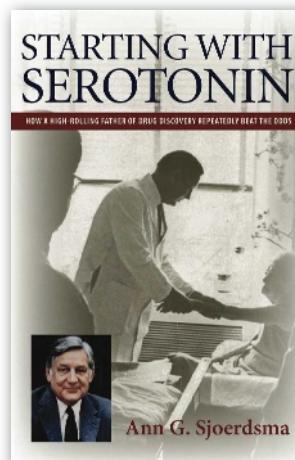
won many devoted friends and colleagues — and lost some, too. But there was no arguing with his successes. The biggest was terfenadine (Seldane), the first antihistamine that did not cause drowsiness. Developed in the 1970s, it was a blockbuster drug until heart arrhythmias surfaced in some users.

Sjoerdsma remained with Richardson-Merrell through the 1980s, until a merger moved the company to Kansas City, Kansas, and gave the management over to a group of businessmen who knew everything about marketing and nothing about science. Sjoerdsma felt the house had changed the rules mid-game. Stripped of his title, and eventually even of his parking space, he felt as if he had lost the shirt off his back.

Ann Sjoerdsma's dual role, as daughter and as professional journalist, creates both windows and blind spots as she examines her father's life. She explains the science and integrates it into Sjoerdsma's career choices and decisions. She draws on interviews she conducted with her father and with his friends and colleagues. The numerous quotations from Sjoerdsma himself, set in italics and without attribution, make it seem as though he is looking over his daughter's shoulder, adding a story or colourful detail, or murmuring assent.

His assent is crucial, for her primary concern is to tell her father's version of his story. In Sjoerdsma's world, the US Food and Drug Administration is a stifling regulatory monster, and it is drugs, more than patients, that live or die. Sometimes he comes across as callous, whereas at other times he is a champion of humanitarian medicine, such as when he developed a therapy for African sleeping sickness. The author rarely questions such views or their motives. Also, she never inquires deeply into Sjoerdsma's emotions.

Particularly striking is the minor role of family in this life portrait. The other Sjoerdsmas feature from time to time, but we never get much sense of how Albert treated them or how



**"Craps was Sjoerdsma's favourite game of chance — and the key theme in his scientific life."**

IMPROBABLE BOOKS

the information necessary to decide whether an object should be taken into the body or expelled.

Many years ago, when working with Gilbert at the Monell Chemical Senses Center in Philadelphia, I recognized his way with words as

well as with scientific research. *What the Nose Knows* melds the academic and business worlds of smell into an entertaining and illuminating rumination on this almost magical sense that, even with a Nobel Prize to its credit, still holds many mysteries. ■

**Gary Beauchamp** is director of the Monell Chemical Senses Center, 3500 Market Street, Philadelphia 19104, Pennsylvania, USA. With Stuart Firestein and David V. Smith, he is co-editor of *Olfaction and Taste*. e-mail: beauchamp@monell.org

## Playing to win

**Starting With Serotonin: How a High-Rolling Father of Drug Discovery Repeatedly Beat the Odds**

by Ann G. Sjoerdsma

Improbable Books: 2008. 640 pp. \$27.50

Science is a gamble. Publication, applying for grants, student admissions and corporate relationships all involve high-stakes bets, a mixture of skill and luck, and often a bit of bluffing. Which game is science most like? It's not a slot machine, mindlessly addictive. In dark moments it may seem like roulette, with its powerful house advantage and long odds. Sometimes it's a horse race, when one thoroughbred laboratory noses out another in isolating a long-sought gene or subatomic particle. For Albert Sjoerdsma, sometimes called the father of clinical pharmacology, science was most like craps.

Craps, an intricate dice game that can involve many players and interweaving rounds of betting, is a thinking person's pastime. Winning depends on an understanding of probability and being able to weigh complex constellations of risks and payouts. In *Starting with Serotonin*, Sjoerdsma's biographer daughter Ann Sjoerdsma argues that craps was her father's favourite game of chance — and the key theme in his scientific life.

Albert Sjoerdsma came to the table with a modest stack of chips. Born in 1924 and raised near Chicago in Illinois, he grew up with little money or social sophistication, but with a first-class brain and a mountain of confidence. He was a rough-and-tumble child, more likely to be found playing sports or getting into mischief than curled up with his nose in a book. Yet his grades were nearly perfect, and he was accepted at the University of Chicago under president Robert Hutchins, whose innovative programmes helped to train some of the best minds of the late twentieth century. Sjoerdsma was cocky and sometimes disrespectful, especially when faced with arbitrary displays of power. His daughter describes him as a maverick, a clichéd but apt term.

Rather than fold with a bachelor's degree, Sjoerdsma stayed in the game, taking an MD and a PhD at Chicago. He then went east to Bethesda, Maryland, where he joined the National Institutes of Health (NIH) in 1951. After two years of residency at the Public Health Service's Marine Hospital in Baltimore, Sjoerdsma landed a position at the National Heart Institute back in Bethesda. There, he formed a team that became known locally as the "wild bunch", a group of brilliant, hard-working and hard-playing researchers.

Sjoerdsma began exploring ways of reducing high blood pressure, leading to his investigation of the then recently discovered neurotransmitter serotonin. His analysis of its effects on different organ systems and metabolic pathways led him into a strongly applied style of pharmacology, in which he largely eschewed lab-based studies in favour of whole patients, and focused on bridging the gap between laboratory science and clinical medicine.

He was a pioneer in the development of monoamine oxidase inhibitors as antidepressants, such as iproniazid, originally developed as an anti-tuberculosis drug. With colleague Sidney Udenfriend, he found that monoamine oxidase was a major pathway for serotonin in both mice and humans, and that inhibitors such as iproniazid raised blood serotonin levels. This established the physiological basis of the antidepressant action of these drugs. Sjoerdsma's

20 years at NIH coincided with the 'golden age' of its intramural research, an era of Nobel prizes, headline-grabbing breakthroughs and major contributions to science.

In 1972, he parlayed his successes as a bench scientist into a job as director of a new research centre in Strasbourg, France, set up by the pharmaceutical company Richardson-Merrell. His blunt, incisive intellectual and administrative style was polarizing in genteel Europe. He

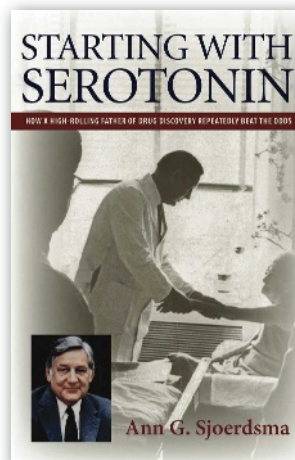
won many devoted friends and colleagues — and lost some, too. But there was no arguing with his successes. The biggest was terfenadine (Seldane), the first antihistamine that did not cause drowsiness. Developed in the 1970s, it was a blockbuster drug until heart arrhythmias surfaced in some users.

Sjoerdsma remained with Richardson-Merrell through the 1980s, until a merger moved the company to Kansas City, Kansas, and gave the management over to a group of businessmen who knew everything about marketing and nothing about science. Sjoerdsma felt the house had changed the rules mid-game. Stripped of his title, and eventually even of his parking space, he felt as if he had lost the shirt off his back.

Ann Sjoerdsma's dual role, as daughter and as professional journalist, creates both windows and blind spots as she examines her father's life. She explains the science and integrates it into Sjoerdsma's career choices and decisions. She draws on interviews she conducted with her father and with his friends and colleagues. The numerous quotations from Sjoerdsma himself, set in italics and without attribution, make it seem as though he is looking over his daughter's shoulder, adding a story or colourful detail, or murmuring assent.

His assent is crucial, for her primary concern is to tell her father's version of his story. In Sjoerdsma's world, the US Food and Drug Administration is a stifling regulatory monster, and it is drugs, more than patients, that live or die. Sometimes he comes across as callous, whereas at other times he is a champion of humanitarian medicine, such as when he developed a therapy for African sleeping sickness. The author rarely questions such views or their motives. Also, she never inquires deeply into Sjoerdsma's emotions.

Particularly striking is the minor role of family in this life portrait. The other Sjoerdsmas feature from time to time, but we never get much sense of how Albert treated them or how



**"Craps was Sjoerdsma's favourite game of chance — and the key theme in his scientific life."**

IMPROBABLE BOOKS

he integrated work and family life. Understandably, the daughter's loyalty seems to trump the journalist's objectivity.

Yet perhaps only she could have played the theme of gambling so well, arguing that her father's strategy was the same at the bench and at the table. He played to win, not to get rich. He was disciplined and used gambling as a test of his willpower. Sjoerdsma's strategy was conservative, but he was willing to take big risks and to bet high stakes when the odds were good. As the author says, he repeatedly beat the odds.

Do not read *Starting With Serotonin* for a story of the selfless thirst for knowledge. Rather, read it for the strategy — both experimental and mercantile — and the passion for competition. Sjoerdsma did it for the love of the game, for better and for worse. ■

**Nathaniel Comfort** is associate professor in the Institute of the History of Medicine at Johns Hopkins University, 1900 East Monument Street, Baltimore 21205, Maryland, USA. His most recent book is *The Panda's Black Box: Opening Up the Intelligent Design Controversy*. e-mail: comfort@jhmi.edu

national government schemes. Phipps argues that progress has been limited by targeting “the problem with girls”, namely that they are gender-stereotyped and fail to recognize the attraction of a career in science. The most successful campaigns described address underlying reasons for girls' career choices, such as the expectations of parents and peer group, and pressures from wider society and culture. Less successful were initiatives that attempted to make science ‘girl-friendly’, for example an after-school club that taught computing skills through topics such as celebrity, fashion and music. As Phipps points out, this widely encountered approach reinforces the stereotypes that such campaigns are seeking to challenge.

Phipps presents statistics to show that universities are among the worst institutions in the United Kingdom in terms of gender equality. Women who choose a career in science may receive support from many professional women's groups and networks. Yet participation is often low owing to limited funding and lack of time to participate. Phipps assesses that the success of initiatives has been limited by focusing on helping women to survive and thrive in existing masculine environments, rather than challenging the underlying work culture and perception of the ideal scientist as a masculine figure: rational, competitive, independent and technically skilled.

If we fail to challenge the interaction between patriarchy and the professionals, Phipps argues, then increasing the numbers of women within technical professions “will not bring about the hoped-for transformation”. In law and medicine, despite the fact that women make up a growing and significant proportion, there remains a marked pay gap, and the working culture is largely unchanged. The book does not provide any quick solutions: increased flexibility in working hours and better maternity pay support a few highly motivated women, usually white and middle-class, to develop their careers. These initiatives also underline society's expectation that women will take on caring, domestic roles as well as professional ones.

We all know talented women who have dropped out of science from a feeling of “not being good enough” to do research or to lead a research group. Phipps's book puts this inner voice to rest by showing that such feelings arise from one's experience of society, culture and capitalism, not one's abilities as a scientist. ■

**Ashleigh Griffin** is a Royal Society Dorothy Hodgkin research fellow at the Institute of Evolutionary Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. e-mail: a.griffin@ed.ac.uk

## Challenging stereotypes

### Women in Science, Engineering and Technology: Three Decades of UK Initiatives

by Alison Phipps

Trentham Books: 2008. 184 pp.

£16.99, €25.50

The lack of women, especially senior women, in science departments is familiar. Less widely appreciated is the effort that has gone into addressing this under-representation. *Women in Science, Engineering and Technology* presents a history of around 150 initiatives to encourage women's participation in science, engineering, construction and technology in the United Kingdom during the past three decades. Alison Phipps, director of gender studies at the University of Sussex, UK, has compiled a valuable resource for activists, policy-makers and educational practitioners, also providing social and political context and analysis.

To a woman working in science, rather than one working for women in science, the book is much more than a reference manual. It puts our experience in context — within the global economy and the women's movement. Phipps makes a compelling case that achieving fundamental change depends on understanding this context and she urges closer communication between educators, academics and social scientists.

Phipps describes how the shift from manufacturing to knowledge-based economies has created demand for skilled workers in science and technology. As in many countries, attracting more women into technical jobs has been identified as a priority for raising Britain's competitive position in the global economy.

Phipps distinguishes between economic and moral motivations to promote gender equality in science, and alerts us to the dangers of relying on the current alignment between the interests of the economy and women's rights. Economic incentives may change with the political climate, leaving women in a vulnerable position if their skills are no longer valued. Also, the ‘business case’ for equality does not concern itself with the underlying reasons for gender inequality, and undermines the idea of equality as an end in itself.

Encouragement for women to pursue careers in science begins in the classroom. *Women in Science, Engineering and Technology* compares grass-roots initiatives in schools with corporate-led and



Schoolgirl Amy's depiction of a typical scientist (left) changed after a visit to a laboratory (right).

he integrated work and family life. Understandably, the daughter's loyalty seems to trump the journalist's objectivity.

Yet perhaps only she could have played the theme of gambling so well, arguing that her father's strategy was the same at the bench and at the table. He played to win, not to get rich. He was disciplined and used gambling as a test of his willpower. Sjoerdsma's strategy was conservative, but he was willing to take big risks and to bet high stakes when the odds were good. As the author says, he repeatedly beat the odds.

Do not read *Starting With Serotonin* for a story of the selfless thirst for knowledge. Rather, read it for the strategy — both experimental and mercantile — and the passion for competition. Sjoerdsma did it for the love of the game, for better and for worse. ■

**Nathaniel Comfort** is associate professor in the Institute of the History of Medicine at Johns Hopkins University, 1900 East Monument Street, Baltimore 21205, Maryland, USA. His most recent book is *The Panda's Black Box: Opening Up the Intelligent Design Controversy*.  
e-mail: comfort@jhmi.edu

national government schemes. Phipps argues that progress has been limited by targeting “the problem with girls”, namely that they are gender-stereotyped and fail to recognize the attraction of a career in science. The most successful campaigns described address underlying reasons for girls' career choices, such as the expectations of parents and peer group, and pressures from wider society and culture. Less successful were initiatives that attempted to make science ‘girl-friendly’, for example an after-school club that taught computing skills through topics such as celebrity, fashion and music. As Phipps points out, this widely encountered approach reinforces the stereotypes that such campaigns are seeking to challenge.

Phipps presents statistics to show that universities are among the worst institutions in the United Kingdom in terms of gender equality. Women who choose a career in science may receive support from many professional women's groups and networks. Yet participation is often low owing to limited funding and lack of time to participate. Phipps assesses that the success of initiatives has been limited by focusing on helping women to survive and thrive in existing masculine environments, rather than challenging the underlying work culture and perception of the ideal scientist as a masculine figure: rational, competitive, independent and technically skilled.

If we fail to challenge the interaction between patriarchy and the professionals, Phipps argues, then increasing the numbers of women within technical professions “will not bring about the hoped-for transformation”. In law and medicine, despite the fact that women make up a growing and significant proportion, there remains a marked pay gap, and the working culture is largely unchanged. The book does not provide any quick solutions: increased flexibility in working hours and better maternity pay support a few highly motivated women, usually white and middle-class, to develop their careers. These initiatives also underline society's expectation that women will take on caring, domestic roles as well as professional ones.

We all know talented women who have dropped out of science from a feeling of “not being good enough” to do research or to lead a research group. Phipps's book puts this inner voice to rest by showing that such feelings arise from one's experience of society, culture and capitalism, not one's abilities as a scientist. ■

**Ashleigh Griffin** is a Royal Society Dorothy Hodgkin research fellow at the Institute of Evolutionary Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK.  
e-mail: a.griffin@ed.ac.uk

## Challenging stereotypes

### Women in Science, Engineering and Technology: Three Decades of UK Initiatives

by Alison Phipps

Trentham Books: 2008. 184 pp.

£16.99, €25.50

The lack of women, especially senior women, in science departments is familiar. Less widely appreciated is the effort that has gone into addressing this under-representation. *Women in Science, Engineering and Technology* presents a history of around 150 initiatives to encourage women's participation in science, engineering, construction and technology in the United Kingdom during the past three decades. Alison Phipps, director of gender studies at the University of Sussex, UK, has compiled a valuable resource for activists, policy-makers and educational practitioners, also providing social and political context and analysis.

To a woman working in science, rather than one working for women in science, the book is much more than a reference manual. It puts our experience in context — within the global economy and the women's movement. Phipps makes a compelling case that achieving fundamental change depends on understanding this context and she urges closer communication between educators, academics and social scientists.

Phipps describes how the shift from manufacturing to knowledge-based economies has created demand for skilled workers in science and technology. As in many countries, attracting more women into technical jobs has been identified as a priority for raising Britain's competitive position in the global economy.

Phipps distinguishes between economic and moral motivations to promote gender equality in science, and alerts us to the dangers of relying on the current alignment between the interests of the economy and women's rights. Economic incentives may change with the political climate, leaving women in a vulnerable position if their skills are no longer valued. Also, the ‘business case’ for equality does not concern itself with the underlying reasons for gender inequality, and undermines the idea of equality as an end in itself.

Encouragement for women to pursue careers in science begins in the classroom. *Women in Science, Engineering and Technology* compares grass-roots initiatives in schools with corporate-led and



Schoolgirl Amy's depiction of a typical scientist (left) changed after a visit to a laboratory (right).

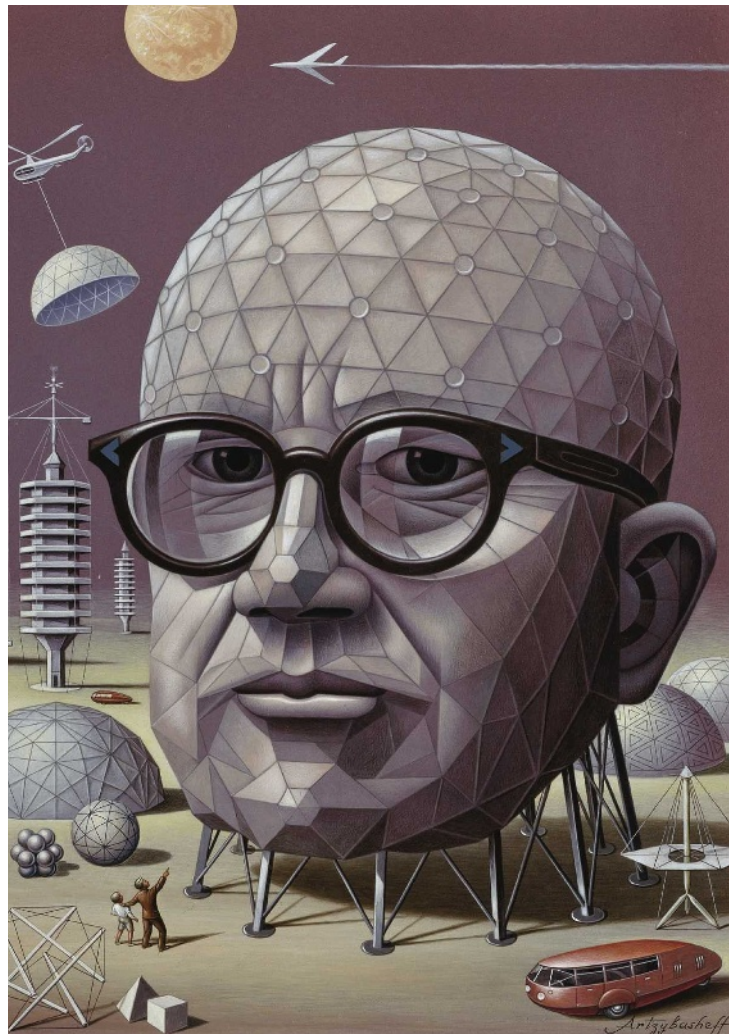
# Bucky's utopian universe

In 1927, on the shore of Lake Michigan, an unemployed and possibly drunk US inventor contemplated suicide. Then, according to the legend he later spun, he had an out-of-body experience. He felt he was floating above the ground, surrounded by celestial light. A voice spoke to him: "You do not have the right to eliminate yourself," it said, "You do not belong to you. You belong to Universe."

Buckminster 'Bucky' Fuller — architect, environmentalist, author and visionary — responded to this call by wholeheartedly attempting to save humanity through science and design. Fifty years of his imaginative and influential work is the subject of a new exhibition at the Whitney Museum of American Art in New York.

Fuller described himself as a "comprehensive, anticipatory design scientist". He is best known for his geodesic dome. Hundreds of them sprung up like mushrooms in the late 1950s and 1960s, housing everything from museums to military bases. Fuller employed tetrahedrons, which he considered to be nature's essential building blocks, to enclose a spherical space using less material than almost any other structure. The Ford Motor Company commissioned one of Fuller's strong and lightweight domes in 1953; six years later, another housed the American National Exhibition in Moscow, the backdrop for the famous 'kitchen debate' between US vice-president Richard Nixon and Soviet premier Nikita Khrushchev. In 1967, a dome formed the centrepiece for the World's Fair in Montreal, Canada. Fuller also planned a dome some three kilometres in diameter to cover mid-town Manhattan like an enormous jellyfish, claiming that it would pay for itself in a decade by the money saved from not having to shovel snow.

The geodesic dome was like a thought bubble



Boris Artzybasheff's 1963 portrait of Buckminster Fuller.

**Buckminster Fuller: Starting With the Universe**

Whitney Museum of American Art, New York  
Until 21 September

encapsulating the utopian dream of modernity. Yet most of Fuller's schemes never left the drawing board. Some of his earliest ideas, on show in the exhibition, are wildly imaginative. He planned a network of airport towers so high that planes could dock above the clouds, imagined planting skyscrapers in craters left behind by Zeppelin-delivered bombs, and in one drawing he attempted to 'synthesize' the Brooklyn Bridge and a Ferris wheel.

Fuller refashioned these outlandish ideas on a domestic scale. In 1928, attempting to solve the US housing crisis, he designed a dwelling whose entire structure hung from a central pole like a merry-go-round, and could

be erected in a day. The buildings looked like spaceships and boasted all the latest appliances, such as air-conditioning and a bathroom unit that required only a litre of water to give you a ten-minute clean with a Fuller-designed fog gun. He persuaded an aircraft firm to build two aluminium prototypes. But the so-called Dymaxion house — a neologism combining 'dynamic', 'maximum' and 'ion' — proved too expensive to mass produce.

Fuller invented a three-wheeled Dymaxion vehicle to complement his "New Era Home". This blimp-like car had a top speed of 190 kilometres an hour, a periscope instead of a rearview mirror, and allowed you to park front-first and then swing into the tightest of spaces. Only three were built — production was halted in 1934 after one crashed outside the entrance to the Chicago World's Fair, killing the driver. The sole surviving example is on view at the Whitney, along with a short film of Fuller showing off the car's tight turning circle.

Fuller's most important legacy is as a futurist who warned of looming environmental catastrophe unless technology was harnessed to create ingenious solutions. If the technology in which he had such faith had

been able to keep pace with his thinking, perhaps more of his unconventional ideas might have been realized.

During Fuller's last decade, the architect Norman Foster collaborated with him, and Foster acknowledges Fuller's influence on his later, environmentally conscious buildings. Off the coast of Dubai, Dutch architects are building maritime structures that echo Fuller's plans for a floating Triton City. Perhaps one day we'll live in the tethered cloud structures that he also dreamed up.

"I did not set out to design a geodesic dome," Fuller wrote. "I set out to discover the principles operative in Universe. For all I knew, this could have led to a pair of flying slippers."

**Christopher Turner** is a writer based in New York and executive editor of *Modern Painters* magazine. e-mail: drchristopherturner@gmail.com

NATIONAL PORTRAIT GALLERY, SMITHSONIAN INSTITUTION; TIME MAGAZINE

# In Retrospect: Gödel's proof

In today's computer age, the implications of the discovery in formal logic that Newman and Nagel articulated in 1958 are of even broader interest, says **Andrew Hodges**.

## Gödel's Proof

by Ernest Nagel and James R. Newman  
New York University Press: 1958. 118 pp.

Fifty years ago an unusual book appeared. Its bald and unapologetic title, *Gödel's Proof*, must have left the casual browser wondering who or what Gödel was. Those tempted to look inside discovered a classic of scientific exposition and faced quite a challenge. The writers, Ernest Nagel and James Newman, already distinguished figures in scientific philosophy and education, gave an uncompromising presentation of their unfamiliar subject matter: mathematical logic. Kurt Gödel himself, the great logician whose breakthrough discovery of 1931 was the subject of this book, was very much alive in 1958 Princeton, but he was no popularizer or media celebrity. For Nagel and Newman to see the potential interest to a wider public was both visionary and optimistic.

I remember discovering *Gödel's Proof* as a student in 1968. I cannot have been the only one to find it a unique text on the college library shelf, leading to unexpected regions beyond the standard syllabus. It was not written like a textbook; neither was it a 'Gödel made easy'. Although rooted in an earlier article in *Scientific American*, it used copious equations; indeed it explored the very meaning of equations, a demand on the reader that would make most publishers nervous. Nagel and Newman explained difficult ideas of logical deduction from formal axioms, distinguishing formal proof from informal reasoning. They showed Gödel's crucial insight: that the rules of logic for quoting axioms, substituting variables and formulating deductions are themselves mathematical operations. And they revealed how his technical innovation exploited this observation, using numbers to code statements about numbers.

Nagel and Newman's detailed account showed how Gödel was led to the astonishing discovery of true mathematical statements that could not possibly have a formal proof. In other words, Gödel proved the formal incompleteness of mathematics. They also recorded the shock that this discovery caused to the hitherto mainstream positivist assumptions, such as those of Bertrand Russell, whose programme for deriving mathematics from purely logical axioms Gödel explicitly contradicted.

The implications of Gödel's discovery are if anything of even broader interest now than

in 1958. A vast industry has arisen founded on logical algorithms, and nowadays it is better appreciated that the business of computing is inseparable from the logical calculus built up in the early twentieth century. One could even argue that the underlying concept of the digital computer is owed to Gödel, via the British mathematician Alan Turing. Turing's 1936 concept of the universal machine is the basis of the computer, and Turing arrived at it by following Gödel's lead, seeing that instructions could operate on other instructions, rather as Gödel's numbers had coded formal statements about numbers.

In the 1950s there was a tendency for mathematicians to distance themselves from practical applications, and from computing in particular. Since the 1970s these divisions have become less rigid. Gödel's arguments are now more fully connected with the body of mathematics and its classical problems of 'how to solve it'. In 2000, the Clay Mathematics Institute in Cambridge, Massachusetts, announced seven prizes for a set of Millennium Problems. One of these, concerning computational complexity, has its root in a remark of Gödel's that might have seemed abstruse in 1958, but is now of great value to practical computing.

It therefore now seems a little odd that Nagel and Newman paid no attention to computing. They framed their closing reflections as if Turing's theory of computability was an obvious corollary. By contrast with their detailed explanation of Gödel's technical arguments, they found no difficulty in writing off, in a few sentences, the possibility of artificial intelligence (AI). This is now a huge and hotly contested area of scientific philosophy. In fact, it was already the subject of dispute in 1958. Turing himself, in 1950, argued that Gödel's proof was irrelevant to the question of achieving AI. In the 1960s, Gödel in turn made somewhat delphic remarks objecting to Turing's philosophy; he seems to have considered that his proof implied that the human mind could not be mechanized.

These arguments stimulated another famous

book with Gödel in the title. Douglas Hofstadter's 1979 *Gödel, Escher, Bach* (Penguin), although in many ways inspired by Newman and Nagel's work, took an approach diametrically opposite to their clipped classicism. Expansive and illustrative, it also came to quite a different conclusion about AI — essentially Turing's. This disagreement remains unresolved; in fact it is heightened by another protagonist, Roger Penrose,

who supports something like Gödel's position but in an entirely new way.

In a few pregnant words, Nagel and Newman referred to the brain as a machine apparently more powerful, through its capacity for informal reasoning, than computers. Penrose, since the 1980s, has asked what could possibly lend it such power, finding an answer in the ill-understood phenomenon of quantum-mechanical state reduction. His conclusions are keenly disputed: for instance the

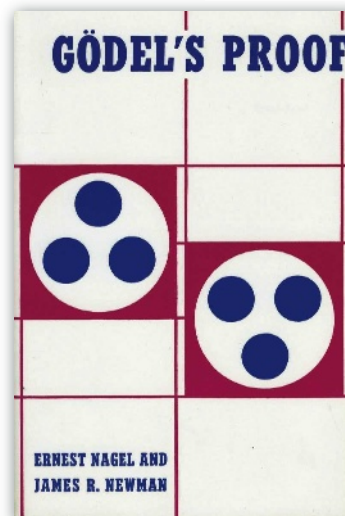
leading logician Martin Davis, himself a popularizer, has forcefully pressed Turing's original view in his book *Engines of Logic* (W. W. Norton, 2001).

Nagel and Newman dedicated their book to Russell, whose logical work during the opening years of the twentieth century lay behind Gödel's proof. In that same period, Planck and Einstein opened the quantum-mechanical door on reality. A hundred years have not sufficed to resolve the fundamental questions they revealed. Mathematical and physical science describe a continuous quantum universe using formal operations on discrete symbols. Neither the quantum, nor those symbols, nor the connection between them, are yet fully understood. ■

**Andrew Hodges** is a Fellow of Wadham College, University of Oxford, Oxford OX1 3PN, UK. He is author of *Alan Turing: the Enigma*. e-mail: andrew.hodges@wadh.ox.ac.uk

## Correction

In Martin Kemp's article on Ferdinand Verbiest (*Nature* 454, 405; 2008), the nationality of Chinese bandit hero Goyo was incorrectly given as Japanese. Also, it is his nickname Chitasei, not the name Wu Yong, that means 'wise star'.



## NEWS &amp; VIEWS



M-SAT LTD/SPL

## QUANTUM MECHANICS

# The speed of instantly

Terence G. Rudolph

**Pairs of quantum-mechanically entangled particles seem to know at once what is happening to each other. Experiments show that even if this signalling is not instantaneous, it must be really, really fast.**

One piece of Einstein's theory of relativity that has taken hold in popular imagination can be summarized by the mantra "nothing travels faster than light". What is less well known is that the theory of quantum mechanics, which deals with the behaviour of very small systems such as atomic and subatomic particles, violates the spirit (if not the letter) of this fundamental principle. Quantum mechanics predicts that, in certain circumstances, an activity performed on one particle can instantaneously change the properties of another particle, no matter how far apart the two particles are. On page 861 of this issue, Salart *et al.*<sup>1</sup> describe an experiment to test how fast 'instantaneous' really is.

One particularly strange feature of quantum mechanics is quantum entanglement. In an experiment involving this phenomenon, a physical property of a particle (or larger system) becomes instantly dependent on the properties that are being measured on another particle, regardless of how far apart the particles are. In a letter to Max Born in 1947, Einstein dismissively called this effect of quantum entanglement a "spooky action at a distance", and thought it indicated that the theory of quantum mechanics was incorrect.

Einstein was not the first to express repugnance at instantaneous action at a distance.

Two hundred and fifty years earlier, Isaac Newton wrote<sup>2</sup>:

... that one body may act upon another at a distance through a vacuum, without the mediation of anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity that I believe no man who has in philosophical matters a competent faculty of thinking can ever fall into it.

Newton was writing about his theory of gravity, and it was Einstein who showed in 1915 that the action of gravity is not instantaneous but is caused by a 'mediation' signal (the warping of space-time), which moves with a finite speed. Salart and colleagues' experiment<sup>1</sup> tests whether the interaction between entangled particles is also conveyed by a mediating signal, and if so, how fast this signal must travel.

In addition to acting instantaneously, and in contrast to other physical effects whose magnitudes vary with distance, the effects of quantum entanglement are predicted to have the same strength no matter how far apart the entangled systems are. Erwin Schrödinger was uncomfortable with this idea, and, in the very papers in which he introduced the term entanglement<sup>3,4</sup>, proposed that some unknown

process must ensure that entanglement occurs over only microscopic distances. He was wrong. The distances over which entanglement has been shown to be maintained increase every year. For example, Salart and colleagues' experiment<sup>1</sup>, which is not designed to test this question, itself measures the entanglement of a pair of photons separated by 18 kilometres.

Salart *et al.* entangled their photon pairs using a source in Geneva, Switzerland, then passed them through fibre-optical cables of exactly equal length to receiving stations in the villages of Jussy and Satigny, which lie respectively east and west of Lake Geneva. Here, the photons' entanglement was checked by an identical pair of interferometers. As they had travelled identical distances, the photons would have reached the interferometers simultaneously, as best as modern optics and electronics allows. Despite this, Salart *et al.* see consistent entanglement of their photons, which means that the time taken by any hypothetical signal passing between them is below the detection limit of the equipment.

There is one subtle feature to all of this, however. Any hypothetical signal has its speed defined in a specific 'preferred frame of reference', which is not the same as the surface of the Earth. Salart *et al.* were able to check their

results against all possible frames of reference by using Earth's rotation. The two villages in which they placed their detectors lie almost exactly east–west of each other, and the authors ran their tests at all hours of the day and night, allowing them to probe every possible orientation of the experiment against a hypothetical preferred reference frame. Taking into account the accuracies of their experimental design and making some conservative assumptions — for example, that Earth is not moving relative to the preferred frame of reference at more than a thousandth the speed of light — Salart *et al.* conclude that any signal passing between the entangled photons is, if not instantaneous, travelling at least ten thousand times faster than light.

The experiment of Salart *et al.*<sup>1</sup> beautifully probes the deep tensions between foundational

aspects of two of our most fundamental physical theories — relativity and quantum mechanics — using quantum entanglement. From it we can conclude that any theory that tries to explain quantum entanglement by invoking a transmission mechanism will need to be very spooky — spookier, perhaps, than quantum mechanics itself.

Terence G. Rudolph is in the Department of Physics, Imperial College London, London SW7 2AZ, UK.

e-mail: t.rudolph@imperial.ac.uk

1. Salart, D., Baas, A., Branciard, C., Gisin, N. & Zbinden, H. *Nature* **454**, 861–864 (2008).
2. Newton, I. in *Isaac Newton: Philosophical Writings* (ed. Janiak, A.) 102 (Cambridge Univ. Press, 2004).
3. Schrödinger, E. *Proc. Camb. Phil. Soc.* **31**, 555–563 (1935).
4. Schrödinger, E. *Proc. Camb. Phil. Soc.* **32**, 446–452 (1936).

## BIOCHEMISTRY

# Fit for an enzyme

Shiven Kapur and Chaitan Khosla

**Certain enzymes that synthesize antibiotics play a game of pass the parcel, handing biosynthetic intermediates from one active site to another. A study reveals the dynamic nature of interactions between the enzyme domains.**

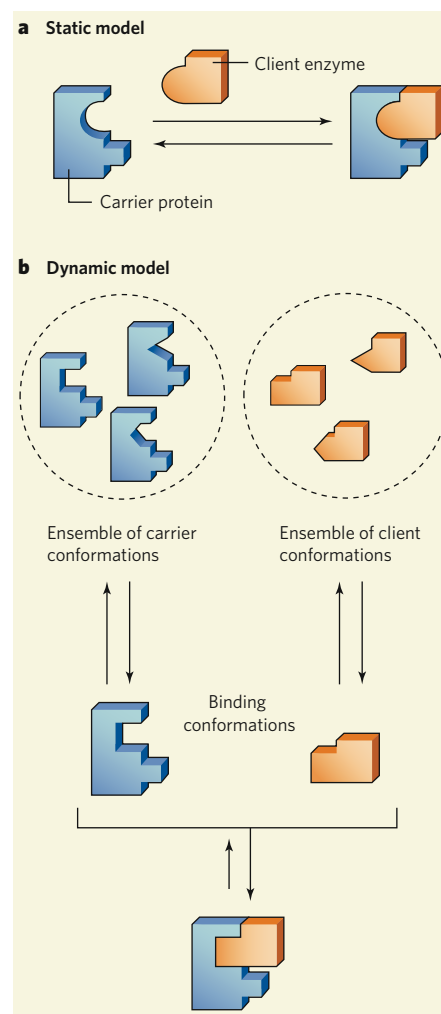
Although crystal structures provide vivid insights into the architecture of enzymes, they reinforce a static picture of the molecules, providing only a snapshot of what a protein looks like in one stable conformation. This can be misleading, because enzymes in solution are certainly not static. Some, such as the megasynthase family of enzymes, are more like molecular versions of factory assembly lines, with moving parts shuttling intermediates from one place to another. Two studies<sup>1,2</sup> in this issue demonstrate the importance of dynamics in the catalytic cycle of a typical megasynthase that makes an antibiotic. Such insight could be used to produce variants of the enzyme that would make analogues of the naturally occurring antibiotic.

The megasynthase family embraces several subclasses of enzyme, but of particular interest to this work are polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), both of which produce antibiotics — including penicillin and vancomycin, the powerful 'drug of last resort'. PKS and NRPS enzymes are large and multifunctional, consisting of clusters of active sites known as modules. Starting with simple organic molecules, each module catalyses a series of reactions in which a new molecular building block is added to the growing chain of the antibiotic, and is then modified to install the chemical groups required in the final product. During this process, biosynthetic intermediates are tethered to carrier proteins, which shuttle them in sequence to designated active sites (known as client enzymes) in the module.

X-ray crystallography has revealed the structures of several PKSs<sup>3</sup> and NRPSs<sup>4</sup>, and those of related enzymes known as fatty-acid synthases<sup>5</sup>. These snapshot pictures provide essential information about their likely mechanisms of action. But there is mounting evidence that protein dynamics has a vital role in enzyme catalysis — for example, the activity of adenylate kinase, a small enzyme often used as a model for biochemical investigations, critically depends on conformational changes that occur on several different timescales<sup>6,7</sup>. It seems likely that dynamic effects are also important in PKS and NRPS enzymes, but this cannot be determined from their crystal structures.

Previous work<sup>8</sup> has revealed the existence of conformational flexibility in the carrier-protein domains of NRPS enzymes when they interact selectively with different client enzymes. Expanding on these findings, Frueh *et al.*<sup>1</sup> (page 903) used nuclear magnetic resonance to determine the structure of a carrier protein in complex with a type I thioesterase — an enzyme that catalyses the final release of antibiotic molecules from PKS and NRPS proteins. The authors observed that part of the thioesterase acts as a lid, flipping between open and closed conformations that expose or conceal the enzyme's carrier-protein binding site. They went on to show that this lid movement is necessary for the tether of the carrier protein to gain access to the active site, thereby catalysing the transfer of the antibiotic to the thioesterase.

Similarly, Koglin *et al.*<sup>2</sup> (page 907) report conformational sub-states of a thioesterase II



**Figure 1 | Protein–protein interactions in megasynthase enzymes.** Many megasynthases make antibiotics. Intermediate compounds in the antibiotic synthetic pathway are transported by carrier proteins to 'client' enzymes for modification. Two models have been proposed to explain the binding of carrier proteins to client enzymes. **a**, In the static model, fixed complementary features in the two proteins are enough to explain carrier–client specificity. **b**, In the dynamic model, specificity arises from the selection of complementary conformations of both the carrier and the client from ensembles of conformations in equilibrium. Non-binding enzymes from the same family as the client aren't recognized by the carrier, because too few of their molecules adopt the correct conformation at equilibrium. Frueh *et al.*<sup>1</sup> and Koglin *et al.*<sup>2</sup> provide evidence for the dynamic model of binding in some megasynthases.

enzyme — a class of protein that repairs the tether of PKS or NRPS enzymes if it becomes 'damaged' through attaching acetyl or related groups. The sub-states all shift towards a single conformation on interacting with an acetylated carrier protein. Taken together with Frueh and colleagues' observations<sup>1</sup>, these results<sup>2</sup> suggest that at least some types of protein–protein interaction in megasynthases are selected from an ensemble of conformations that exist in dynamic equilibrium with each other. Such

results against all possible frames of reference by using Earth's rotation. The two villages in which they placed their detectors lie almost exactly east–west of each other, and the authors ran their tests at all hours of the day and night, allowing them to probe every possible orientation of the experiment against a hypothetical preferred reference frame. Taking into account the accuracies of their experimental design and making some conservative assumptions — for example, that Earth is not moving relative to the preferred frame of reference at more than a thousandth the speed of light — Salart *et al.* conclude that any signal passing between the entangled photons is, if not instantaneous, travelling at least ten thousand times faster than light.

The experiment of Salart *et al.*<sup>1</sup> beautifully probes the deep tensions between foundational

aspects of two of our most fundamental physical theories — relativity and quantum mechanics — using quantum entanglement. From it we can conclude that any theory that tries to explain quantum entanglement by invoking a transmission mechanism will need to be very spooky — spookier, perhaps, than quantum mechanics itself.

Terence G. Rudolph is in the Department of Physics, Imperial College London, London SW7 2AZ, UK.

e-mail: t.rudolph@imperial.ac.uk

1. Salart, D., Baas, A., Branciard, C., Gisin, N. & Zbinden, H. *Nature* **454**, 861–864 (2008).
2. Newton, I. in *Isaac Newton: Philosophical Writings* (ed. Janiak, A.) 102 (Cambridge Univ. Press, 2004).
3. Schrödinger, E. *Proc. Camb. Phil. Soc.* **31**, 555–563 (1935).
4. Schrödinger, E. *Proc. Camb. Phil. Soc.* **32**, 446–452 (1936).

## BIOCHEMISTRY

# Fit for an enzyme

Shiven Kapur and Chaitan Khosla

**Certain enzymes that synthesize antibiotics play a game of pass the parcel, handing biosynthetic intermediates from one active site to another. A study reveals the dynamic nature of interactions between the enzyme domains.**

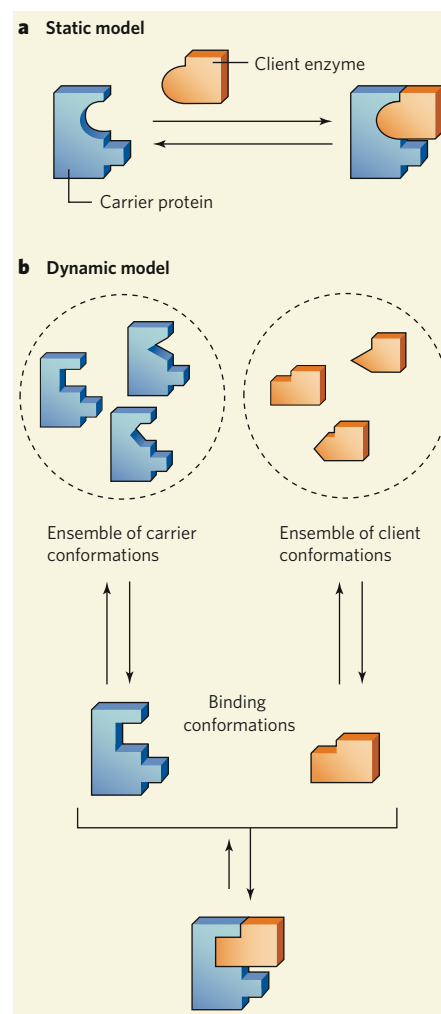
Although crystal structures provide vivid insights into the architecture of enzymes, they reinforce a static picture of the molecules, providing only a snapshot of what a protein looks like in one stable conformation. This can be misleading, because enzymes in solution are certainly not static. Some, such as the megasynthase family of enzymes, are more like molecular versions of factory assembly lines, with moving parts shuttling intermediates from one place to another. Two studies<sup>1,2</sup> in this issue demonstrate the importance of dynamics in the catalytic cycle of a typical megasynthase that makes an antibiotic. Such insight could be used to produce variants of the enzyme that would make analogues of the naturally occurring antibiotic.

The megasynthase family embraces several subclasses of enzyme, but of particular interest to this work are polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs), both of which produce antibiotics — including penicillin and vancomycin, the powerful 'drug of last resort'. PKS and NRPS enzymes are large and multifunctional, consisting of clusters of active sites known as modules. Starting with simple organic molecules, each module catalyses a series of reactions in which a new molecular building block is added to the growing chain of the antibiotic, and is then modified to install the chemical groups required in the final product. During this process, biosynthetic intermediates are tethered to carrier proteins, which shuttle them in sequence to designated active sites (known as client enzymes) in the module.

X-ray crystallography has revealed the structures of several PKSs<sup>3</sup> and NRPSs<sup>4</sup>, and those of related enzymes known as fatty-acid synthases<sup>5</sup>. These snapshot pictures provide essential information about their likely mechanisms of action. But there is mounting evidence that protein dynamics has a vital role in enzyme catalysis — for example, the activity of adenylate kinase, a small enzyme often used as a model for biochemical investigations, critically depends on conformational changes that occur on several different timescales<sup>6,7</sup>. It seems likely that dynamic effects are also important in PKS and NRPS enzymes, but this cannot be determined from their crystal structures.

Previous work<sup>8</sup> has revealed the existence of conformational flexibility in the carrier-protein domains of NRPS enzymes when they interact selectively with different client enzymes. Expanding on these findings, Frueh *et al.*<sup>1</sup> (page 903) used nuclear magnetic resonance to determine the structure of a carrier protein in complex with a type I thioesterase — an enzyme that catalyses the final release of antibiotic molecules from PKS and NRPS proteins. The authors observed that part of the thioesterase acts as a lid, flipping between open and closed conformations that expose or conceal the enzyme's carrier-protein binding site. They went on to show that this lid movement is necessary for the tether of the carrier protein to gain access to the active site, thereby catalysing the transfer of the antibiotic to the thioesterase.

Similarly, Koglin *et al.*<sup>2</sup> (page 907) report conformational sub-states of a thioesterase II



**Figure 1 | Protein–protein interactions in megasynthase enzymes.** Many megasynthases make antibiotics. Intermediate compounds in the antibiotic synthetic pathway are transported by carrier proteins to 'client' enzymes for modification. Two models have been proposed to explain the binding of carrier proteins to client enzymes. **a**, In the static model, fixed complementary features in the two proteins are enough to explain carrier–client specificity. **b**, In the dynamic model, specificity arises from the selection of complementary conformations of both the carrier and the client from ensembles of conformations in equilibrium. Non-binding enzymes from the same family as the client aren't recognized by the carrier, because too few of their molecules adopt the correct conformation at equilibrium. Frueh *et al.*<sup>1</sup> and Koglin *et al.*<sup>2</sup> provide evidence for the dynamic model of binding in some megasynthases.

enzyme — a class of protein that repairs the tether of PKS or NRPS enzymes if it becomes 'damaged' through attaching acetyl or related groups. The sub-states all shift towards a single conformation on interacting with an acetylated carrier protein. Taken together with Frueh and colleagues' observations<sup>1</sup>, these results<sup>2</sup> suggest that at least some types of protein–protein interaction in megasynthases are selected from an ensemble of conformations that exist in dynamic equilibrium with each other. Such

a model has previously been proposed<sup>9,10</sup> to explain certain protein–protein interactions.

Now that protein dynamics has been shown to play a role in antibiotic biosynthesis by NRPS (and thus probably also PKS) enzymes, the next step is to understand which, if any, of the many dynamic modes are involved in controlling the specificity of interactions between proteins. The importance of specific protein–protein interactions in such biosyntheses is well established<sup>11,12</sup>, but a better understanding of them is essential if the enzymes are to be rationally re-engineered to synthesize new antibiotics. Thus far, attention has focused mainly on identifying static features that allow protein domains to dock selectively to each other, such as regions that have complementary shapes or electric charges (Fig. 1a). The assumption is that such features will be similar for evolutionarily related client enzymes, but will be different for other enzyme families. If so, one should be able to replace any individual client enzyme with a related enzyme, without perturbing the rest of the megasynthase.

But if protein–protein specificity is the result of favourable docking between selected equilibrium conformations of the proteins' domains (Fig. 1b), then rational design of megasynthases requires knowledge of an entirely different set of issues. For example, how similar are the conformational distributions of analogous domains in related proteins? Are the energy barriers for conformational switching comparable for analogous domains across an entire family of enzymes? And does selective docking always require the same pair of compatible conformations?

The principles of physical organic chemistry were founded on the study of small molecules, and have provided a strong basis for elucidating the roles of static features in protein–protein recognition. But these ideas are of limited use in determining the rules that govern protein dynamics. Time will tell whether these dynamics are pivotal to the behaviour of megasynthases. If we are ever to be able to modify such enzymes to prepare new antibiotics, we can only hope that evolution has been parsimonious in writing the design rules. ■

Shiven Kapur and Chaitan Khosla are in the Department of Chemistry, Stanford University, Stanford, California 94305, USA.  
e-mails: shiven@stanford.edu;  
khosla@stanford.edu

1. Frueh, D. P. *et al.* *Nature* **454**, 903–906 (2008).
2. Koglin, A. *et al.* *Nature* **454**, 907–911 (2008).
3. Tang, Y., Kim, C.-Y., Mathews, I. L., Cane, D. E. & Khosla, C. *Proc. Natl Acad. Sci. USA* **103**, 11124–11129 (2006).
4. Tanovic, A. *et al.* *Science* **321**, 659–663 (2008).
5. Jenni, S. *et al.* *Science* **316**, 254–261 (2007).
6. Wolf-Watz, M. *et al.* *Nature Struct. Mol. Biol.* **11**, 945–949 (2004).
7. Henzler-Wildman, K. A. *et al.* *Nature* **450**, 913–916 (2007).
8. Koglin, A. *et al.* *Science* **312**, 273–276 (2006).
9. Kumar, S. *et al.* *Protein Sci.* **9**, 10–19 (2000).
10. Grünberg, R., Leckner, J. & Nilges, M. *Structure* **12**, 2125–2136 (2004).
11. Khosla, C. *et al.* *Annu. Rev. Biochem.* **76**, 195–221 (2007).
12. Lai, J. R., Koglin, A. & Walsh, C. T. *Biochemistry* **45**, 14869–14879 (2006).

## EARTH SCIENCE

# Solid evidence in the inner core

Kenneth C. Creager

**The indications are that a solid ball of iron lies at Earth's centre. But only the identification of an elusive seismic signature can confirm the long-standing assumption that it is indeed solid.**

It is widely accepted that Earth's inner core is solid. Seismologists have long been trying to find direct evidence of this assertion — which could come in the form of the identification of a particular kind of seismic wave, called a shear wave, travelling through the inner core. Perhaps, as reported by Wookey and Helffrich on page 873 of this issue<sup>1</sup>, that elusive wave has now been teased out of the background seismic noise.

Earth is made up of a solid crust and mantle overlying outer (fluid) and inner cores that consist primarily of iron. Large earthquakes send two types of seismic wave — compressional elastic and shear elastic waves — right through Earth. Both kinds of wave travel through a solid, but only compressional waves travel through a fluid. Thus, unambiguous detection of a shear wave travelling through the inner core would be once-and-for-all evidence that it is solid. The inner-core shear wave is extremely elusive because of its tortuous journey through Earth's centre; only a smidgen of energy remains after it converts between a compressional elastic wave in the fluid outer core and a shear elastic wave in the inner core on its transit both into and then back out of the inner core.

Seismologists label waves in the mantle and crust as P for compression or S for shear; those in the outer core as K for compression; and those in the inner core as I for compression or J for shear. Wookey and Helffrich<sup>1</sup> analysed PKJKP. In other words, they looked for a compressional wave emitted from an earthquake that travelled down through the crust, mantle and outer core (PK); that became a shear wave in the inner core (J); and that then travelled back up to the surface as a compressional wave (KP) where it could be recorded.

The inner core was discovered about 70 years ago by Inge Lehmann<sup>2</sup>. She used the timing of elastic compressional waves to show that the speed of these waves increases abruptly at the boundary of the inner core. Francis Birch<sup>3</sup> interpreted this increase as 'freezing' of the iron, resulting in a solid inner core and a fluid outer core. Near the centre of Earth, temperature increases slowly with depth, whereas pressure increases rapidly. So even though the inner core is hotter than the outer core, Birch inferred that it is solid because of the very high pressure. The observed increase in compressional-wave speed going into the inner core is compatible with that predicted for a fluid–solid transition. Although Birch's interpretation that

the inner core is solid is widely accepted, it is not definitive. Hence the quest for better evidence.

Large earthquakes cause Earth to resonate like a bell at a set of distinct frequencies, and studies of some of these modes has provided evidence for the solidity of the inner core. However, this kind of analysis is complicated, and it is difficult to prove beyond a doubt that no model with a fluid inner core could explain the results. Which brings us back to the gold standard for evidence of a solid inner core — a compelling observation of shear waves passing through it. Indeed, there are four previous reports of observing a shear wave that went through the inner core, two at high frequency<sup>4,5</sup> (above 0.1 Hz) and two at low frequency<sup>6,7</sup> (below 0.1 Hz). The lower-frequency observations are arguably more robust, but all such observations are challenging because these waves are similar in amplitude to many others that scatter off irregularities inside Earth.

Wookey and Helffrich<sup>1</sup> analysed more than 700 high-quality seismograms from the Japanese Hi-net array of borehole seismometers (15–100 times as many stations as those used in the previous studies). Processing data from this array as if it were a downward-looking telescope allowed the authors to determine the precise timing and direction of seismic waves. They observed one wave from one earthquake that arrived at almost the exact time and direction required for it to be an inner-core shear wave as predicted by a standard model of wave velocity known as ak135 (ref. 8). This was followed, 7 seconds later, by another arrival, which they interpreted as a second shear wave that had been split by elastic anisotropy in the inner core. In an anisotropic medium, a shear wave is split into two shear waves with orthogonal polarizations that travel at different wave speeds. The inner core is known to be anisotropic from the many analyses of the directional dependence of the speed of compressional waves. But this is the first reported observation of shear-wave splitting in the inner core.

A surprising result of Wookey and Helffrich's analysis is the relatively high frequencies (0.2–0.4 Hz) at which the shear wave is observed. Because the inner core is near its melting temperature, it is expected to be strongly attenuating. Attenuation is characterized by a quality factor, *Q*, which is proportional to the number of wave cycles required for energy to decay by a factor of 1/*e*. Most existing models predict that attenuation would cause PKJKP to be vanishingly small at (and

a model has previously been proposed<sup>9,10</sup> to explain certain protein–protein interactions.

Now that protein dynamics has been shown to play a role in antibiotic biosynthesis by NRPS (and thus probably also PKS) enzymes, the next step is to understand which, if any, of the many dynamic modes are involved in controlling the specificity of interactions between proteins. The importance of specific protein–protein interactions in such biosyntheses is well established<sup>11,12</sup>, but a better understanding of them is essential if the enzymes are to be rationally re-engineered to synthesize new antibiotics. Thus far, attention has focused mainly on identifying static features that allow protein domains to dock selectively to each other, such as regions that have complementary shapes or electric charges (Fig. 1a). The assumption is that such features will be similar for evolutionarily related client enzymes, but will be different for other enzyme families. If so, one should be able to replace any individual client enzyme with a related enzyme, without perturbing the rest of the megasynthase.

But if protein–protein specificity is the result of favourable docking between selected equilibrium conformations of the proteins' domains (Fig. 1b), then rational design of megasynthases requires knowledge of an entirely different set of issues. For example, how similar are the conformational distributions of analogous domains in related proteins? Are the energy barriers for conformational switching comparable for analogous domains across an entire family of enzymes? And does selective docking always require the same pair of compatible conformations?

The principles of physical organic chemistry were founded on the study of small molecules, and have provided a strong basis for elucidating the roles of static features in protein–protein recognition. But these ideas are of limited use in determining the rules that govern protein dynamics. Time will tell whether these dynamics are pivotal to the behaviour of megasynthases. If we are ever to be able to modify such enzymes to prepare new antibiotics, we can only hope that evolution has been parsimonious in writing the design rules. ■

Shiven Kapur and Chaitan Khosla are in the Department of Chemistry, Stanford University, Stanford, California 94305, USA.  
e-mails: shiven@stanford.edu;  
khosla@stanford.edu

1. Frueh, D. P. *et al.* *Nature* **454**, 903–906 (2008).
2. Koglin, A. *et al.* *Nature* **454**, 907–911 (2008).
3. Tang, Y., Kim, C.-Y., Mathews, I. L., Cane, D. E. & Khosla, C. *Proc. Natl Acad. Sci. USA* **103**, 11124–11129 (2006).
4. Tanovic, A. *et al.* *Science* **321**, 659–663 (2008).
5. Jenni, S. *et al.* *Science* **316**, 254–261 (2007).
6. Wolf-Watz, M. *et al.* *Nature Struct. Mol. Biol.* **11**, 945–949 (2004).
7. Henzler-Wildman, K. A. *et al.* *Nature* **450**, 913–916 (2007).
8. Koglin, A. *et al.* *Science* **312**, 273–276 (2006).
9. Kumar, S. *et al.* *Protein Sci.* **9**, 10–19 (2000).
10. Grünberg, R., Leckner, J. & Nilges, M. *Structure* **12**, 2125–2136 (2004).
11. Khosla, C. *et al.* *Annu. Rev. Biochem.* **76**, 195–221 (2007).
12. Lai, J. R., Koglin, A. & Walsh, C. T. *Biochemistry* **45**, 14869–14879 (2006).

## EARTH SCIENCE

# Solid evidence in the inner core

Kenneth C. Creager

**The indications are that a solid ball of iron lies at Earth's centre. But only the identification of an elusive seismic signature can confirm the long-standing assumption that it is indeed solid.**

It is widely accepted that Earth's inner core is solid. Seismologists have long been trying to find direct evidence of this assertion — which could come in the form of the identification of a particular kind of seismic wave, called a shear wave, travelling through the inner core. Perhaps, as reported by Wookey and Helffrich on page 873 of this issue<sup>1</sup>, that elusive wave has now been teased out of the background seismic noise.

Earth is made up of a solid crust and mantle overlying outer (fluid) and inner cores that consist primarily of iron. Large earthquakes send two types of seismic wave — compressional elastic and shear elastic waves — right through Earth. Both kinds of wave travel through a solid, but only compressional waves travel through a fluid. Thus, unambiguous detection of a shear wave travelling through the inner core would be once-and-for-all evidence that it is solid. The inner-core shear wave is extremely elusive because of its tortuous journey through Earth's centre; only a smidgen of energy remains after it converts between a compressional elastic wave in the fluid outer core and a shear elastic wave in the inner core on its transit both into and then back out of the inner core.

Seismologists label waves in the mantle and crust as P for compression or S for shear; those in the outer core as K for compression; and those in the inner core as I for compression or J for shear. Wookey and Helffrich<sup>1</sup> analysed PKJKP. In other words, they looked for a compressional wave emitted from an earthquake that travelled down through the crust, mantle and outer core (PK); that became a shear wave in the inner core (J); and that then travelled back up to the surface as a compressional wave (KP) where it could be recorded.

The inner core was discovered about 70 years ago by Inge Lehmann<sup>2</sup>. She used the timing of elastic compressional waves to show that the speed of these waves increases abruptly at the boundary of the inner core. Francis Birch<sup>3</sup> interpreted this increase as 'freezing' of the iron, resulting in a solid inner core and a fluid outer core. Near the centre of Earth, temperature increases slowly with depth, whereas pressure increases rapidly. So even though the inner core is hotter than the outer core, Birch inferred that it is solid because of the very high pressure. The observed increase in compressional-wave speed going into the inner core is compatible with that predicted for a fluid–solid transition. Although Birch's interpretation that

the inner core is solid is widely accepted, it is not definitive. Hence the quest for better evidence.

Large earthquakes cause Earth to resonate like a bell at a set of distinct frequencies, and studies of some of these modes has provided evidence for the solidity of the inner core. However, this kind of analysis is complicated, and it is difficult to prove beyond a doubt that no model with a fluid inner core could explain the results. Which brings us back to the gold standard for evidence of a solid inner core — a compelling observation of shear waves passing through it. Indeed, there are four previous reports of observing a shear wave that went through the inner core, two at high frequency<sup>4,5</sup> (above 0.1 Hz) and two at low frequency<sup>6,7</sup> (below 0.1 Hz). The lower-frequency observations are arguably more robust, but all such observations are challenging because these waves are similar in amplitude to many others that scatter off irregularities inside Earth.

Wookey and Helffrich<sup>1</sup> analysed more than 700 high-quality seismograms from the Japanese Hi-net array of borehole seismometers (15–100 times as many stations as those used in the previous studies). Processing data from this array as if it were a downward-looking telescope allowed the authors to determine the precise timing and direction of seismic waves. They observed one wave from one earthquake that arrived at almost the exact time and direction required for it to be an inner-core shear wave as predicted by a standard model of wave velocity known as ak135 (ref. 8). This was followed, 7 seconds later, by another arrival, which they interpreted as a second shear wave that had been split by elastic anisotropy in the inner core. In an anisotropic medium, a shear wave is split into two shear waves with orthogonal polarizations that travel at different wave speeds. The inner core is known to be anisotropic from the many analyses of the directional dependence of the speed of compressional waves. But this is the first reported observation of shear-wave splitting in the inner core.

A surprising result of Wookey and Helffrich's analysis is the relatively high frequencies (0.2–0.4 Hz) at which the shear wave is observed. Because the inner core is near its melting temperature, it is expected to be strongly attenuating. Attenuation is characterized by a quality factor, *Q*, which is proportional to the number of wave cycles required for energy to decay by a factor of 1/*e*. Most existing models predict that attenuation would cause PKJKP to be vanishingly small at (and

above) 0.3 Hz. As the authors point out, their observation may require changes in our ideas about the inner-core attenuation: perhaps the strongly attenuating part of the inner core is concentrated in the upper 400 km; perhaps  $Q$  for shear waves is much larger (meaning less attenuation) at 0.3 Hz than its measured value at lower frequencies; or perhaps attenuation in pure compression is comparable to that in pure shear.

The inner core is its own world-within-a-world; separated from the mantle and crust by a fluid outer core, it can behave somewhat independently. It seems to be spinning slightly faster than the rest of the planet. In addition, wave speeds in the inner core depend strongly on the direction in which the waves travel, implying that much of the inner core is

composed of strongly aligned crystals. This alignment is apparent at scales from 1 km — over which it causes waves to scatter — to a hemispheric scale, in which compressional waves travelling parallel to the spin axis in the western hemisphere consistently produce some of the largest high-frequency compressional-wave travel-time anomalies on Earth.

Much about the inner core remains shrouded from view — for example, it is not clear whether crystals are aligned as the inner core cools and grows, or by slow shear flow at a later stage. Wookey and Helffrich<sup>1</sup> may well have opened a new door into the darkness. Whatever the case, the future of this type of analysis is bright — facilities such as Hi-net, and another seismic-sensing system, the EarthScope Transportable Array, offer new

possibilities for exploring the inner workings of Earth's inner core, as well as other parts of our planet.

Kenneth C. Creager is in the Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195-1310, USA.  
e-mail: kcc@ess.washington.edu

1. Wookey, J. & Helffrich, G. *Nature* **454**, 873–876 (2008).
2. Lehmann, I. *Bur. Centr. Seismol. Int. A* **14**, 87–115 (1936).
3. Birch, F. *Am. J. Sci.* **238**, 192–211 (1940).
4. Julian, B. R., Davies, D. & Sheppard, R. M. *Nature* **235**, 317–318 (1972).
5. Okal, E. A. & Cansi, Y. *Earth Planet. Sci. Lett.* **164**, 23–30 (1998).
6. Deuss, A., Woodhouse, J. H., Paulssen, H. & Trampert, J. *Geophys. J. Int.* **142**, 67–73 (2000).
7. Cao, A., Romanowicz, B. & Takeuchi, N. *Science* **308**, 1453–1455 (2005).
8. Kennett, B. L. N., Engdahl, E. R. & Buland, R. *Geophys. J. Int.* **122**, 108–124 (1995).

## PALAEONTOLOGY

# The last giant kangaroo

Jared Diamond

**Humans who colonized Australia did not reach Tasmania until thousands of years later — granting the island's giant kangaroos a brief respite before they joined their Australian brethren in oblivion.**

Many of the world's biggest animals (the megafauna) became extinct in the latter part of the last ice age, which ended around 14,000 years ago. That realization triggered two long-standing debates<sup>1</sup>. First, were the extinctions caused by humans, climatic or vegetational changes, fires, extraterrestrial impacts, or some combination of those agents? Second, if humans were responsible, did they cause the extinctions directly through hunting, or indirectly by some consequences of human settlement? Turney *et al.*<sup>2</sup> now contribute to answering both of these questions. Writing in *Proceedings of the National Academy of Sciences*, they report improved dating of the demise of giant kangaroos and other giant marsupials, and of associated changes (or lack of them), on the island of Tasmania, off Australia.

Megafaunal extinction proceeded to different extents and at different times on different continents. If one defines megafauna as species with an adult body mass of more than 44 kilograms, then most megafaunal species living in the latter part of the last ice age have survived until today in Africa and Eurasia. But most or all of them became extinct in Australia about 46,000 years ago, in North America about 13,000 years ago, in South America about 13,000 years ago and in New Zealand about 700 years ago. Those four dates are close to the respective dates of human entry into each of these land masses, suggesting that humans were somehow responsible.

The first humans to colonize these four regions were anatomically and behaviourally modern humans with sophisticated hunting

skills, who would have encountered animals with no previous experience of humans. Those tame animals would have paid the price for their lack of fear of these sudden invaders — in much the same way, during recent centuries, that animals of Antarctica and the Galapagos initially had no fear of humans. In contrast, the animals of Africa and Eurasia were saved by their gradual acquisition of fear of humans over millions of years, as human hunting ability slowly developed from modest ancestral skills such as those of chimpanzees. Although

the coincidences in time point the finger at humans as the cause of megafaunal extinctions, a complication is that extinctions in North and South America coincided approximately not only with human arrival there, but also with a peak of global warming and glacial retreat at the end of the last ice age — as well as (in North America) with a sediment layer suggestive of an extraterrestrial impact<sup>3</sup>.

Arguments against a direct human role in the extinction of Australia's megafauna stem from reports of vegetational changes and increased fires on the Australian mainland around the same time<sup>4,5</sup>, and especially from the evidence for Tasmania available until now<sup>6,7</sup>. Today, Tasmania is an island that lies 240 kilometres southeast of the Australian mainland. But the seas between Australia and Tasmania are so shallow that Tasmania was intermittently connected to Australia during glacial episodes, when sea level dropped because much of the world's water was locked up in ice. By the time humans colonized Australia around



**Figure 1 | Two extinct species of giant kangaroo.** The tall animal is *Protomnodon anak*, fossils of which were excavated by Turney *et al.*<sup>2</sup> at the Mount Cripps site in Tasmania. This species, which stood about 2 metres tall and weighed 100–150 kilograms, survived on Tasmania until the arrival of humans. The other species depicted is *Protomnodon tumbuna*, which formerly lived in the mountains of New Guinea and was the only species of kangaroo that didn't hop. (Painting by Peter Schouten.)

above) 0.3 Hz. As the authors point out, their observation may require changes in our ideas about the inner-core attenuation: perhaps the strongly attenuating part of the inner core is concentrated in the upper 400 km; perhaps  $Q$  for shear waves is much larger (meaning less attenuation) at 0.3 Hz than its measured value at lower frequencies; or perhaps attenuation in pure compression is comparable to that in pure shear.

The inner core is its own world-within-a-world; separated from the mantle and crust by a fluid outer core, it can behave somewhat independently. It seems to be spinning slightly faster than the rest of the planet. In addition, wave speeds in the inner core depend strongly on the direction in which the waves travel, implying that much of the inner core is

composed of strongly aligned crystals. This alignment is apparent at scales from 1 km — over which it causes waves to scatter — to a hemispheric scale, in which compressional waves travelling parallel to the spin axis in the western hemisphere consistently produce some of the largest high-frequency compressional-wave travel-time anomalies on Earth.

Much about the inner core remains shrouded from view — for example, it is not clear whether crystals are aligned as the inner core cools and grows, or by slow shear flow at a later stage. Wookey and Helffrich<sup>1</sup> may well have opened a new door into the darkness. Whatever the case, the future of this type of analysis is bright — facilities such as Hi-net, and another seismic-sensing system, the EarthScope Transportable Array, offer new

possibilities for exploring the inner workings of Earth's inner core, as well as other parts of our planet.

Kenneth C. Creager is in the Department of Earth and Space Sciences, University of Washington, Seattle, Washington 98195-1310, USA.  
e-mail: kcc@ess.washington.edu

1. Wookey, J. & Helffrich, G. *Nature* **454**, 873–876 (2008).
2. Lehmann, I. *Bur. Centr. Seismol. Int. A* **14**, 87–115 (1936).
3. Birch, F. *Am. J. Sci.* **238**, 192–211 (1940).
4. Julian, B. R., Davies, D. & Sheppard, R. M. *Nature* **235**, 317–318 (1972).
5. Okal, E. A. & Cansi, Y. *Earth Planet. Sci. Lett.* **164**, 23–30 (1998).
6. Deuss, A., Woodhouse, J. H., Paulssen, H. & Trampert, J. *Geophys. J. Int.* **142**, 67–73 (2000).
7. Cao, A., Romanowicz, B. & Takeuchi, N. *Science* **308**, 1453–1455 (2005).
8. Kennett, B. L. N., Engdahl, E. R. & Buland, R. *Geophys. J. Int.* **122**, 108–124 (1995).

## PALAEONTOLOGY

# The last giant kangaroo

Jared Diamond

**Humans who colonized Australia did not reach Tasmania until thousands of years later — granting the island's giant kangaroos a brief respite before they joined their Australian brethren in oblivion.**

Many of the world's biggest animals (the megafauna) became extinct in the latter part of the last ice age, which ended around 14,000 years ago. That realization triggered two long-standing debates<sup>1</sup>. First, were the extinctions caused by humans, climatic or vegetational changes, fires, extraterrestrial impacts, or some combination of those agents? Second, if humans were responsible, did they cause the extinctions directly through hunting, or indirectly by some consequences of human settlement? Turney *et al.*<sup>2</sup> now contribute to answering both of these questions. Writing in *Proceedings of the National Academy of Sciences*, they report improved dating of the demise of giant kangaroos and other giant marsupials, and of associated changes (or lack of them), on the island of Tasmania, off Australia.

Megafaunal extinction proceeded to different extents and at different times on different continents. If one defines megafauna as species with an adult body mass of more than 44 kilograms, then most megafaunal species living in the latter part of the last ice age have survived until today in Africa and Eurasia. But most or all of them became extinct in Australia about 46,000 years ago, in North America about 13,000 years ago, in South America about 13,000 years ago and in New Zealand about 700 years ago. Those four dates are close to the respective dates of human entry into each of these land masses, suggesting that humans were somehow responsible.

The first humans to colonize these four regions were anatomically and behaviourally modern humans with sophisticated hunting

skills, who would have encountered animals with no previous experience of humans. Those tame animals would have paid the price for their lack of fear of these sudden invaders — in much the same way, during recent centuries, that animals of Antarctica and the Galapagos initially had no fear of humans. In contrast, the animals of Africa and Eurasia were saved by their gradual acquisition of fear of humans over millions of years, as human hunting ability slowly developed from modest ancestral skills such as those of chimpanzees. Although

the coincidences in time point the finger at humans as the cause of megafaunal extinctions, a complication is that extinctions in North and South America coincided approximately not only with human arrival there, but also with a peak of global warming and glacial retreat at the end of the last ice age — as well as (in North America) with a sediment layer suggestive of an extraterrestrial impact<sup>3</sup>.

Arguments against a direct human role in the extinction of Australia's megafauna stem from reports of vegetational changes and increased fires on the Australian mainland around the same time<sup>4,5</sup>, and especially from the evidence for Tasmania available until now<sup>6,7</sup>. Today, Tasmania is an island that lies 240 kilometres southeast of the Australian mainland. But the seas between Australia and Tasmania are so shallow that Tasmania was intermittently connected to Australia during glacial episodes, when sea level dropped because much of the world's water was locked up in ice. By the time humans colonized Australia around



**Figure 1 | Two extinct species of giant kangaroo.** The tall animal is *Protomnodon anak*, fossils of which were excavated by Turney *et al.*<sup>2</sup> at the Mount Cripps site in Tasmania. This species, which stood about 2 metres tall and weighed 100–150 kilograms, survived on Tasmania until the arrival of humans. The other species depicted is *Protomnodon tumbuna*, which formerly lived in the mountains of New Guinea and was the only species of kangaroo that didn't hop. (Painting by Peter Schouten.)

50,000 years ago, Tasmania was an island, and it did not become reconnected to Australia and accessible to human settlement until 43,000 years ago. Hence, claims that megafaunal species were already extinct in Tasmania before humans arrived, if supported, would be fatal to the hypothesis that humans eliminated the megafauna in Tasmania, and, by extension, in Australia.

Turney *et al.*<sup>2</sup> have now compared the Tasmanian chronologies of megafaunal extinction and human arrival, and lake-core evidence of climate and vegetational changes and of fire charcoal, thereby providing a record unique in the Australian region for correlating all of those indicators within a small area. The authors reanalysed museum specimens of fossils from four previously reported sites, as well as assessing finds from a newly excavated site at Mount Cripps in the northwest of the island.

The overall Tasmanian record extends from the present back to at least 127,000 years ago, during the last interglacial period. It turns out that, as expected, there were drastic climatic and vegetational changes (cool temperate rainforest yielding to grassland, herbs and heath) around 110,000 years ago, at the transition from the last interglacial to the last glacial period. Those changes were reversed after the transition from glacial back to interglacial conditions around 15,000 years ago. But the switch to glacial climates and vegetation 110,000 years ago had no effect on the megafauna's composition. All seven species recorded on Tasmania in the last interglacial, more than 127,000 years ago, were still present in the most extensive glacial fossil deposit, Scotchtown Cave, dated at 56,000 years ago — strong evidence against the cause of the extinctions being climatic and vegetational changes before humans arrived. The seven now-vanished giants consisted of two 500-kilogram marsupial equivalents of a rhinoceros and a ground sloth; three antelope-like kangaroos weighing 100–150 kilograms and standing 2 metres tall; a leopard-like marsupial predator; and a giant echidna (not a marsupial but a monotreme).

The site at Mount Cripps yielded three specimens of the giant kangaroo, *Protemnodon anak* (Fig. 1), dated at between 42,900 and 40,900 years ago, an age statistically identical to the age of the oldest archaeological horizon (excavation level) in Tasmania, as dated by carbon-14, that contains evidence of human occupation. This discovery supports two conclusions. First, despite Tasmania's proximity and vegetational similarity to southeastern Australia, megafauna survived later in Tasmania than on the Australian mainland, from which they had disappeared about 46,000 years ago. Second, in Tasmania the megafauna overlapped in time with humans, who probably arrived when the land bridge was reestablished 43,000 years ago, as indicated by undated human cultural artefacts lying below the level of the carbon-14-dated human horizon. Thus, megafauna that had survived in Tasmania for at least

84,000 years (from 127,000 to 43,000 years ago), through big changes in vegetation and climate, disappeared within a few millennia of human arrival.

The results of Turney *et al.*<sup>2</sup> also point to the cause of that event. Unlike the Australian mainland, Tasmania shows no signs of changes in vegetation or in fire charcoal around the time of human arrival. Those changes on the mainland have been interpreted<sup>4,5</sup> to mean that humans eliminated the mainland megafauna indirectly, by lighting fires that altered the character of the vegetation. The absence of such evidence in Tasmania suggests that humans there instead eliminated megafauna directly by hunting. Even for the mainland, the role of fire-induced vegetational changes as a cause of megafaunal extinction is debated: vegetational changes might have been the result, rather than the cause, of the elimination of large browsing and grazing mammals.

Obviously, our understanding of the chronology and cause of the Tasmanian megafaunal extinction is still far from complete. As sceptics will point out, Turney *et al.* have demonstrated the survival of one species of giant kangaroo in Tasmania until humans arrived. What about the other six species of megafauna?

I, too, hope for more complete evidence. Proof to convince any sceptic would be the existence of many Tasmanian sites containing human remains dated at 42,800 years ago, all lying above deposits lacking human material and thus indicating human arrival at that date, plus remains of all seven megafaunal species at that same date, followed by the last remains of those species at seven different dates ranging from 42,700 to 41,900 years ago. That's asking a lot. But Turney and colleagues' study may motivate Tasmanian palaeontologists and archaeologists to seek such overwhelming evidence, or else to refute that scenario. ■

Jared Diamond is in the Department of Geography, University of California, Los Angeles, California 90095-1524, USA.

e-mail: jdiamond@geog.ucla.edu

1. Martin, P. S. & Klein, R. G. *Quaternary Extinctions: A Prehistoric Revolution* (Univ. Arizona Press, 1984).
2. Turney, C. S. M. *et al.* *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0801360105 (2008).
3. Haynes, C. V. Jr *Proc. Natl Acad. Sci. USA* **105**, 6520–6525 (2008).
4. Miller, J. H. *et al.* *Science* **309**, 287–290 (2008).
5. Lynch, A. H. *et al.* *Annu. Rev. Earth Planet. Sci.* **35**, 215–239 (2007).
6. Wroe, S. & Field, J. *Quat. Sci. Rev.* **25**, 2692–2703 (2006).
7. Cosgrove, R. in *Encyclopedia of Quaternary Science* (ed. Elias, S. A.) 118–129 (Elsevier, 2007).

## NANOTECHNOLOGY

# Shaping the void

Astrid Lambrecht

**A vacuum may be devoid of matter, but its shape is still important. The strength of the Casimir force caused by quantum fluctuations in the space between surfaces is critically dependent on their nanometre-scale shape.**

Two reflecting plates facing each other in a vacuum attract. This startling phenomenon is known as the Casimir effect<sup>1</sup>, named after the Dutch physicist Hendrik Casimir who discovered it in 1948. It is a quantum-mechanical effect that arises because even a perfect vacuum is not entirely empty but filled with unavoidable fluctuations of the electromagnetic field, imposed by the Heisenberg uncertainty principle, and imbuing the vacuum with a non-zero energy. The reflecting plates provide boundaries between which only some electromagnetic fluctuations can exist compared with those in the unrestricted vacuum outside the plates. The vacuum energy between the plates is thus different from outside, resulting in the Casimir 'force', which scales with the fourth power of the inverse distance between the plates and is proportional to their surface area, making it, strictly speaking, a pressure. Writing in *Physical Review Letters*, Chan *et al.*<sup>2</sup> find that the precise shape of the plates also has an effect by measuring the Casimir force between a gold-covered sphere and a nanostructured silicon surface.

Casimir forces are relatively long-range in comparison to atomic forces, but only become measurable at sub-micrometre scales. In the mesoscopic regime (~100 nanometres) they are quite substantial and have been measured with increasing precision during the past decade. In their experiment, Chan *et al.*<sup>2</sup> used a micro-oscillator, a device introduced for Casimir-force measurements some years ago<sup>3,4</sup>. It consists of a silicon surface, suspended in such a way that it can be set oscillating by electrical excitation. Two gold-coated glass spheres of radius ~50 µm are stacked on top of each other and attached to one side of the silicon plate. In isolation, this system undergoes harmonic oscillations at its resonance frequency, but when a surface is brought close to the topmost sphere, the Casimir attraction between them alters the oscillator's resonance frequency. Chan *et al.* measured this change using different nanostructured silicon surfaces containing repeating patterns of square-sided grooves around 1 µm deep, with a periodicity of either 400 nm or 1 µm, and brought within 150–500 nm of the gold sphere. The torsional

50,000 years ago, Tasmania was an island, and it did not become reconnected to Australia and accessible to human settlement until 43,000 years ago. Hence, claims that megafaunal species were already extinct in Tasmania before humans arrived, if supported, would be fatal to the hypothesis that humans eliminated the megafauna in Tasmania, and, by extension, in Australia.

Turney *et al.*<sup>2</sup> have now compared the Tasmanian chronologies of megafaunal extinction and human arrival, and lake-core evidence of climate and vegetational changes and of fire charcoal, thereby providing a record unique in the Australian region for correlating all of those indicators within a small area. The authors reanalysed museum specimens of fossils from four previously reported sites, as well as assessing finds from a newly excavated site at Mount Cripps in the northwest of the island.

The overall Tasmanian record extends from the present back to at least 127,000 years ago, during the last interglacial period. It turns out that, as expected, there were drastic climatic and vegetational changes (cool temperate rainforest yielding to grassland, herbs and heath) around 110,000 years ago, at the transition from the last interglacial to the last glacial period. Those changes were reversed after the transition from glacial back to interglacial conditions around 15,000 years ago. But the switch to glacial climates and vegetation 110,000 years ago had no effect on the megafauna's composition. All seven species recorded on Tasmania in the last interglacial, more than 127,000 years ago, were still present in the most extensive glacial fossil deposit, Scotchtown Cave, dated at 56,000 years ago — strong evidence against the cause of the extinctions being climatic and vegetational changes before humans arrived. The seven now-vanished giants consisted of two 500-kilogram marsupial equivalents of a rhinoceros and a ground sloth; three antelope-like kangaroos weighing 100–150 kilograms and standing 2 metres tall; a leopard-like marsupial predator; and a giant echidna (not a marsupial but a monotreme).

The site at Mount Cripps yielded three specimens of the giant kangaroo, *Protemnodon anak* (Fig. 1), dated at between 42,900 and 40,900 years ago, an age statistically identical to the age of the oldest archaeological horizon (excavation level) in Tasmania, as dated by carbon-14, that contains evidence of human occupation. This discovery supports two conclusions. First, despite Tasmania's proximity and vegetational similarity to southeastern Australia, megafauna survived later in Tasmania than on the Australian mainland, from which they had disappeared about 46,000 years ago. Second, in Tasmania the megafauna overlapped in time with humans, who probably arrived when the land bridge was reestablished 43,000 years ago, as indicated by undated human cultural artefacts lying below the level of the carbon-14-dated human horizon. Thus, megafauna that had survived in Tasmania for at least

84,000 years (from 127,000 to 43,000 years ago), through big changes in vegetation and climate, disappeared within a few millennia of human arrival.

The results of Turney *et al.*<sup>2</sup> also point to the cause of that event. Unlike the Australian mainland, Tasmania shows no signs of changes in vegetation or in fire charcoal around the time of human arrival. Those changes on the mainland have been interpreted<sup>4,5</sup> to mean that humans eliminated the mainland megafauna indirectly, by lighting fires that altered the character of the vegetation. The absence of such evidence in Tasmania suggests that humans there instead eliminated megafauna directly by hunting. Even for the mainland, the role of fire-induced vegetational changes as a cause of megafaunal extinction is debated: vegetational changes might have been the result, rather than the cause, of the elimination of large browsing and grazing mammals.

Obviously, our understanding of the chronology and cause of the Tasmanian megafaunal extinction is still far from complete. As sceptics will point out, Turney *et al.* have demonstrated the survival of one species of giant kangaroo in Tasmania until humans arrived. What about the other six species of megafauna?

I, too, hope for more complete evidence. Proof to convince any sceptic would be the existence of many Tasmanian sites containing human remains dated at 42,800 years ago, all lying above deposits lacking human material and thus indicating human arrival at that date, plus remains of all seven megafaunal species at that same date, followed by the last remains of those species at seven different dates ranging from 42,700 to 41,900 years ago. That's asking a lot. But Turney and colleagues' study may motivate Tasmanian palaeontologists and archaeologists to seek such overwhelming evidence, or else to refute that scenario. ■

Jared Diamond is in the Department of Geography, University of California, Los Angeles, California 90095-1524, USA.

e-mail: jdiamond@geog.ucla.edu

1. Martin, P. S. & Klein, R. G. *Quaternary Extinctions: A Prehistoric Revolution* (Univ. Arizona Press, 1984).
2. Turney, C. S. M. *et al.* *Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0801360105 (2008).
3. Haynes, C. V. Jr *Proc. Natl Acad. Sci. USA* **105**, 6520–6525 (2008).
4. Miller, J. H. *et al.* *Science* **309**, 287–290 (2008).
5. Lynch, A. H. *et al.* *Annu. Rev. Earth Planet. Sci.* **35**, 215–239 (2007).
6. Wroe, S. & Field, J. *Quat. Sci. Rev.* **25**, 2692–2703 (2006).
7. Cosgrove, R. in *Encyclopedia of Quaternary Science* (ed. Elias, S. A.) 118–129 (Elsevier, 2007).

## NANOTECHNOLOGY

# Shaping the void

Astrid Lambrecht

**A vacuum may be devoid of matter, but its shape is still important. The strength of the Casimir force caused by quantum fluctuations in the space between surfaces is critically dependent on their nanometre-scale shape.**

Two reflecting plates facing each other in a vacuum attract. This startling phenomenon is known as the Casimir effect<sup>1</sup>, named after the Dutch physicist Hendrik Casimir who discovered it in 1948. It is a quantum-mechanical effect that arises because even a perfect vacuum is not entirely empty but filled with unavoidable fluctuations of the electromagnetic field, imposed by the Heisenberg uncertainty principle, and imbuing the vacuum with a non-zero energy. The reflecting plates provide boundaries between which only some electromagnetic fluctuations can exist compared with those in the unrestricted vacuum outside the plates. The vacuum energy between the plates is thus different from outside, resulting in the Casimir 'force', which scales with the fourth power of the inverse distance between the plates and is proportional to their surface area, making it, strictly speaking, a pressure. Writing in *Physical Review Letters*, Chan *et al.*<sup>2</sup> find that the precise shape of the plates also has an effect by measuring the Casimir force between a gold-covered sphere and a nanostructured silicon surface.

Casimir forces are relatively long-range in comparison to atomic forces, but only become measurable at sub-micrometre scales. In the mesoscopic regime (~100 nanometres) they are quite substantial and have been measured with increasing precision during the past decade. In their experiment, Chan *et al.*<sup>2</sup> used a micro-oscillator, a device introduced for Casimir-force measurements some years ago<sup>3,4</sup>. It consists of a silicon surface, suspended in such a way that it can be set oscillating by electrical excitation. Two gold-coated glass spheres of radius ~50 µm are stacked on top of each other and attached to one side of the silicon plate. In isolation, this system undergoes harmonic oscillations at its resonance frequency, but when a surface is brought close to the topmost sphere, the Casimir attraction between them alters the oscillator's resonance frequency. Chan *et al.* measured this change using different nanostructured silicon surfaces containing repeating patterns of square-sided grooves around 1 µm deep, with a periodicity of either 400 nm or 1 µm, and brought within 150–500 nm of the gold sphere. The torsional

axis of the oscillator was arranged perpendicularly to the grooves' orientation in order to avoid motion due to a lateral Casimir force arising between corrugated surfaces<sup>5</sup>. The measured Casimir forces were smaller than those produced between spheres and smooth surfaces by a factor varying between 1.5 and 1.8, depending on the distance and surface structure.

The reduction in force can be explained by the simple argument that the top parts of the surface alone are close enough to interact efficiently with the sphere. However, the tops of the surface constitute half of the area of a flat surface and would be expected to produce half the Casimir force. This reasoning has been the basis of a widely used simplification known as the 'proximity force approximation'. In order to find the Casimir force between bodies with more complicated shapes than simple planes, the proximity force approximation splits up the shaped surfaces into (infinitely) small planar surfaces and sums the Casimir forces between these surface elements as if they were independent of each other. The force between a sphere and a flat plate then becomes the integral of the pressure over the surface of the sphere. This approximation has been tested experimentally<sup>6</sup> and works so well that it has been used extensively to evaluate the Casimir force for numerous experiments in which, for practical reasons, the force has been measured between a sphere and a plate<sup>7,8</sup>, and not in the original Casimir configuration of two parallel plates.

The force between the nanostructured silicon plate and the sphere in the experiment by Chan *et al.*<sup>2</sup> is greater than would be predicted by the proximity force approximation. This provides the first experimental evidence that Casimir forces are not in general additive, despite being weak forces. Explicit theoretical calculations for the force in this configuration predict a difference twice as large as the measured discrepancy if the boundaries are considered to be perfect reflectors<sup>9</sup>, but come closer to the experimental result when the finite reflectivity of gold and silicon is taken into account<sup>10</sup>.

Why does the proximity force approximation correctly predict the Casimir force between a sphere and a flat surface, but not for nanostructured surfaces? As often in physics, it is a matter of differing length scales — even vacuum energy is not immune to this rule. The distance between the two boundaries (two plates, plate and sphere, or sphere and nanostructured surface) constitutes a characteristic length scale of the system. Vacuum fluctuations are reflected from both boundaries, and those fluctuations with wavelengths similar to the separation of the boundaries contribute most to the Casimir force. The spheres used to measure Casimir forces have radii of about 50  $\mu\text{m}$ , much larger than the plate–sphere separations of a few hundred nanometres over which the Casimir force is detectable. The curves of these spheres are

on a length scale too large to have a noticeable effect on the reflection process and the Casimir force. But the surfaces used by Chan *et al.*<sup>2</sup> have crenellations with periods of 400 nm and 1  $\mu\text{m}$ , on the same scale as the distance over which the Casimir force operates. Their precise shape thus has a substantial influence on the reflection of the most relevant electromagnetic fluctuations and so alters the strength of the Casimir force<sup>11</sup>.

Chan *et al.*<sup>2</sup> have shown that nanostructures give shape to vacuum energy in a detectable way, even if micro- or macrostructures do not. It now remains to be seen whether an appropriate choice of materials and surface structures can tune the Casimir force 'at will' for nanotechnological applications. ■

Astrid Lambrecht is at the Laboratoire Kastler Brossel, Campus Jussieu, Case 74, 75252 Paris Cedex 05, France.

1. Casimir, H. B. G. *Proc. K. Ned. Akad. Wet.* **51**, 793–795 (1948).
2. Chan, H. B. *et al. Phys. Rev. Lett.* **101**, 030401 (2008).
3. Buks, E. & Roukes, M. L. *Europhys. Lett.* **54**, 220–226 (2001).
4. Chan, H. B., Aksyuk, V. A., Kleiman, R. N., Bishop, D. J. & Capasso, F. *Science* **291**, 1941–1944 (2001).
5. Chen, F., Mohideen, U., Klimchitskaya, G. L. & Mostepanenko, V. M. *Phys. Rev. Lett.* **88**, 101801 (2002).
6. Krause, D. E., Decca, R. S., López, D. & Fischbach, E. *Phys. Rev. Lett.* **98**, 050403 (2007).
7. Lamoreaux, S. K. *Phys. Rev. Lett.* **78**, 5–8 (1997).
8. Mohideen, U. & Roy, A. *Phys. Rev. Lett.* **81**, 4549–4552 (1998).
9. Büscher, R. & Emig, T. *Phys. Rev. A* **69**, 062101 (2004).
10. Lambrecht, A. & Marachevsky, V. N. preprint at <http://arxiv.org/abs/0806.3142> (2008).
11. Rodrigues, R. B., Maia Neto, P. A., Lambrecht, A. & Reynaud, S. *Phys. Rev. A* **75**, 062108 (2007).

## IMMUNOLOGY

# Surprising side effects

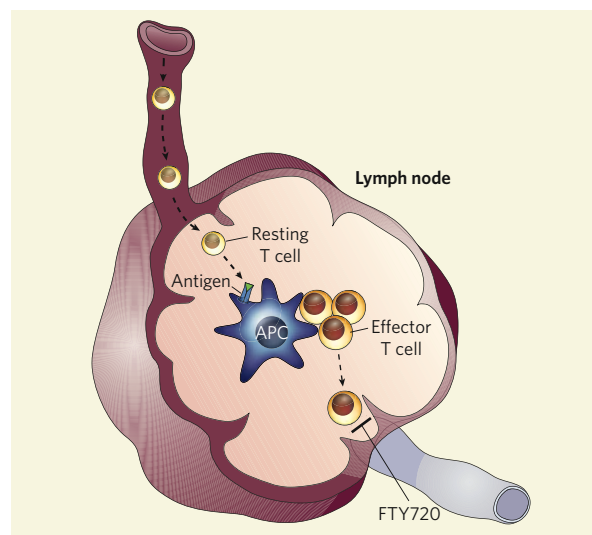
Michael J. Bevan and Pamela J. Fink

**A drug that normally suppresses an immune response by trapping lymphocytes in lymphoid organs results in the elimination of a chronic viral infection when applied at low doses. Why should this be?**

During many immune responses, the circulation of blood lymphocytes, such as T cells, to lymph nodes and back to the bloodstream via lymphatic vessels is temporarily shut off. Lymphocytes can still enter the nodes, but their egress is curtailed, leading to a dramatic decline in the number of these cells in the blood — a condition called lymphopenia. On page 894 of this issue, Premenko-Lanier *et al.*<sup>1</sup> report that administration of a drug that enhances such lymphocyte trapping converts what would be a chronic viral infection into one that is rapidly cleared by the immune system. The observation gives an unexpected twist to the potential side effects of this immunosuppressive drug.

Two closely related strains of a mouse pathogen — lymphocytic choriomeningitis virus — differ dramatically in their course of infection. The parent strain, known as Armstrong, is rapidly cleared by potent killer T cells and an accompanying antibody response. The derivative strain, clone 13, which differs from Armstrong by only two amino-acid residues, likewise induces a strong immune response, but the virus manages to escape this response and infected mice retain high viral titres in their blood and tissues for many months. With time, high levels of viral antigen reduce or exhaust the functionality of antiviral T cells, including their ability to kill infected cells and

**Figure 1 | Blocked lymphocyte recirculation.** During an immune response, T cells recirculate from the blood into the lymph nodes and back into the blood again. On encountering antigens on an antigen-presenting cell (APC), T cells that have entered the lymph node from the blood rapidly divide and differentiate into effector cells before re-entering the blood. Premenko-Lanier *et al.*<sup>1</sup> find that FTY720, which blocks lymphocyte egress from the lymph nodes, somehow aids the immune response to infection with the clone 13 strain of lymphocytic choriomeningitis virus.



secrete immune-enhancing molecules called cytokines<sup>2</sup>.

The impetus for the current work<sup>1</sup> came when the researchers noticed that, although acute infection with Armstrong leads to a sharp decline in lymphocyte numbers in the blood through efficient trapping in the lymphoid organs, infection with clone 13 virus did not cause lymphopenia. Could the failure to eliminate clone 13 be due to this difference in lymphocyte circulation?

Lymphocyte egress from lymph nodes is controlled by signals from sphingosine-1-phosphate (S1P), which acts on G-protein-coupled cell-surface receptors. A new class of immunosuppressive pro-drug, typified by FTY720, also signals via these receptors, blocking lymphocyte egress from lymph nodes<sup>3,4</sup> (Fig. 1). Premenko-Lanier *et al.*<sup>1</sup> find that administering low doses of this drug during the first few days of clone 13 infection leads to efficient lymphocyte trapping in the nodes and clearance of the virus in seven days. Even more remarkably, short-term drug treatment of mice that have been persistently infected with clone 13 virus for 30 days — and consequently whose effector T cells are showing signs of exhaustion — leads to viral clearance.

Although the rationale for applying S1P analogues was based on the observation that clone 13 infection did not lead to lymphocyte trapping, it is not clear that this is the explanation for the drug's effect. There are five known S1P receptors, and these have wide tissue distribution. FTY720 reacts with four of these, influencing not only lymphocyte recirculation, but also other phenomena such as heart rate and vascular integrity<sup>3,4</sup>. Whether the critical effect of the drug is on lymphocytes themselves, endothelial cells or another cell type is not known. Future analysis of this phenomenon will probably use newer ligands that have defined target specificity for S1P receptors. It is clear, however, that drug treatment does not directly abolish infection in some unsuspected way, as results show that drug-induced clearance requires an intact immune system<sup>1</sup>.

Why does clone 13 usually escape the strong immune response? Again, the answer is not definitely known but may relate to the fact that clone 13 has a wider tissue distribution than the parent Armstrong virus. It infects more cells in the lymph node, such as antigen-presenting cells<sup>5</sup> and reticular cells<sup>6</sup>, and its explosive growth rate may simply overwhelm the nascent immune response. One side effect of the more widespread infection by clone 13 may be a disruption of the normal lymphocyte-trapping mechanism. It is reasonable to propose that enhancing lymphocyte trapping in lymph nodes augments the initial response to clone 13 infection. In support of this argument, the absence of lymph nodes alters the course of what would be an acute Armstrong infection into a chronic infection<sup>7</sup>.

It is harder to imagine how drug treatment leads to clearance of an established infection. The authors do not state whether the drug

regimen applied at the chronic stage of clone 13 infection caused lymphopenia by enhanced lymphocyte retention in the lymph nodes. At this late time point, lymphocytes specific for viral antigens are already showing signs of exhaustion and must be rejuvenated. It seems likely that FTY720 boosts the response, either by enhancing positive regulators of immunity or by blocking negative regulators. CD4-type T cells are necessary for FTY720's effects on clone 13 infection, and these cells are crucial for reversing the dysfunction in antiviral killer T cells. Whatever the mechanism for the antiviral effect of FTY720 treatment, Premenko-Lanier and colleagues' work<sup>1</sup> serves to put clinicians who prescribe this drug on notice to expect some unusual side effects. ■

Michael J. Bevan and Pamela J. Fink are in the Department of Immunology, University of Washington, Seattle, Washington 98195-7370, USA.

e-mails: mbevan@u.washington.edu; pfink@u.washington.edu

1. Premenko-Lanier, M., Moseley, N. B., Pruett, S. T., Romagnoli, P. A. & Altman, J. D. *Nature* **454**, 894–898 (2008).
2. Wherry, E. J., Blattman, J. N., Murali-Krishna, K., van der Most, R. & Ahmed, R. *J. Virol.* **77**, 4911–4927 (2003).
3. Schwab, S. R. & Cyster, J. G. *Nature Immunol.* **8**, 1295–1301 (2007).
4. Rosen, H. *et al. Immunol. Rev.* **223**, 221–235 (2008).
5. Borrow, P., Evans, C. F. & Oldstone, M. B. *J. Virol.* **69**, 1059–1070 (1995).
6. Mueller, S. N. *et al. Proc. Natl Acad. Sci. USA* **104**, 15430–15435 (2007).
7. Karrer, U. *et al. J. Exp. Med.* **185**, 2157–2170 (1997).

## AUTISM

# Family connections

Daniel H. Geschwind

**Autism is a common neurodevelopmental syndrome with a strong genetic component. The study of autistic individuals whose parents are cousins highlights the genetic diversity of this condition.**

Genetically complex human disorders, such as obesity and Alzheimer's disease, also occur in rare inherited forms — that is, a disease-causing mutation arises that can be passed on to the next generation. Studying such rare mutations is a powerful strategy for gaining insight into the mechanisms of the more common forms of a disease<sup>1</sup>. This is especially true of autism, for which most associated gene mutations that have been identified are either rare or cause syndromic forms of the disorder<sup>2–6</sup>. Reporting in *Science*, Morrow *et al.*<sup>7</sup> describe their study of a large group of highly consanguineous families — families in which there were many marriages between cousins — to search for genes connected with increased susceptibility to autism.

To identify rare, disease-causing, recessively inherited mutations in consanguineous families, a technique called homozygosity mapping<sup>8</sup> is used. In this technique, the genomes of a group of subjects with a certain disorder are searched for regions that they have in common and that are identical (homozygous) on both chromosomes of an affected individual. It is assumed that such regions are passed on from both parents, who shared a recent common ancestor. Homozygosity mapping has been successfully used to identify the genes responsible for certain inherited diseases, such as rare disorders of neural development, but it has not been widely applied to more common, complex genetic disorders.

Morrow *et al.*<sup>7</sup> aimed to carry out homozygosity mapping to identify autism genes in children of consanguineous marriages. But

consanguineous families in which several children have autism are hard to come by, and so the first challenge was to find and characterize such pedigrees.

The authors therefore established an impressive collaborative network throughout the Middle East and Eurasia (the Homozygosity Mapping Collaborative for Autism; HMCA), and identified 88 consanguineous families with autistic children. (It is common for cousins in Middle Eastern families to marry, and families tend to be large, so the search for inherited factors is that much easier.) The authors scanned the genomes of their subjects at high resolution for structural chromosomal anomalies such as deletions or duplications (also called copy-number variations, or CNVs).

Previous work<sup>3,9</sup> has found that individuals with sporadic (non-familial) autism have a higher prevalence of *de novo* CNVs (those not observed in parents) than do those with inherited autism. Morrow *et al.* do not find such an increase in *de novo* CNVs in their subjects, but they do find an increase in inherited CNVs in HMCA families. This, along with a low level of *de novo* events, is consistent with inherited, rather than sporadic, autism in these families.

The authors also identify six chromosomal regions in the HMCA subjects that harbour inherited, homozygous deletions ranging from 18 to 880 kilobases in size. The largest of these regions removes a gene of unknown function called *c3orf58*, and disrupts the promoter sequence of another gene, *NHE9*, which encodes a cell-membrane protein. The authors

compared the sequence of *NHE9* in several hundred autistic subjects from non-consanguineous families and controls, and show that single base-pair mutations in the gene are associated with autism. Their finding therefore further supports a role for *NHE9* in increasing the susceptibility to this disorder.

The symptoms of autism include communication and social deficits, and restricted and repetitive behaviour. Clinical features can vary widely between individuals, and may include epilepsy or developmental regression after a period of apparently normal development. To simplify their analysis of *NHE9*, Morrow and colleagues focused on a subset of individuals who suffered from autism-associated epilepsy. It is thought that a subset of individuals who also have a particular additional disease feature (phenotype) in common might form a more genetically similar group. Unlike most of the autistic subjects without epilepsy, these individuals showed a more than fivefold increase in potentially harmful base-pair changes in the *NHE9* gene sequence, including a stop codon. Therefore, *NHE9* may be added to the list of genes that are associated with certain subsets of autism<sup>6</sup>.

Morrow and co-workers' observations<sup>7</sup> raise some intriguing questions. Five of the six homozygous deletions identified in the HMCA subjects lie in regions outside the coding sequence of the implicated genes, and so are unlikely to disrupt the encoded protein sequence. Several of the deletions might disrupt presumptive regulatory regions, whereas another 320-kilobase deletion closest to the *PCDH10* gene, which encodes the PCD10 (protocadherin 10) protein, does not lie in a coding region or in a sequence of known function. The effect of these deletions on neighbouring genes, and any causal role in autism, therefore require further investigation. Also, the causal nature of the *NHE9* mutations, which result in potentially deleterious changes at the protein level, must be established, because similar changes are observed in control subjects, albeit less frequently.

The difficulty in identifying mutations that are definitely pathogenic is an emerging theme in common neuropsychiatric disorders. Morrow and colleagues' study demonstrates that studying consanguineous families is a fruitful path to identifying autism genes, but also that such analysis is more challenging than might initially have been expected. This is partly due to the underlying genetic heterogeneity of autism, a facet underscored by this study. The authors find that no single recessive mutation is associated with autism in more than one family, even in consanguineous families from the same geographic region.

Luckily, advances in DNA sequencing technology mean that even large genomic regions can be efficiently sequenced, and work to identify recessive mutations in single base pairs within the genomes of subjects from the HMCA group might already be under way.

Sequencing shared regions in these families could be even more informative because many mutations in these regions are probably small — perhaps single base-pair mutations — and thus might more directly implicate specific genes.

Convergent evidence from other studies, including functional and gene-expression data, will be crucial in assessing the relevance of candidate genes to autism<sup>10</sup>. Morrow *et al.* embrace this notion, as their putative autism susceptibility genes represent diverse functional classes, which at face value do not suggest common molecular mechanisms. To provide a potential functional context, they show that three of the genes — either near (*PCDH10*) or overlapping (*c3orf58* and *NHE9*) the deletions — are among a significant subset that is associated with a form of neuronal activity thought to be involved in learning and memory<sup>11</sup>.

The authors suggest that investigating such genes might provide a way of unifying seemingly diverse molecular findings in autism. Given the nearly two dozen genes that are now associated with the disorder<sup>6</sup>, this should be a readily testable hypothesis. The way in which functional data relate to the core deficits of autism is not known, because many aspects of learning and memory are preserved in autistic children. The idea that autism might represent a general deficit in neuronal plasticity caused by changes in gene expression needs to be reconciled with the specificity of cognitive

dysfunction observed in the condition<sup>12</sup>, as well as the preservation of general intellectual ability in many individuals. This aspect highlights a central challenge of research in this area: connecting basic molecular pathways to the properties of complex neural circuits and cognition in humans. Spanning these diverse levels of function will require detailed analysis of the correlation between the genetic composition of individuals and their associated phenotype across the spectrum of autism disorders.

Daniel H. Geschwind is in the Program in Neurogenetics, Department of Neurology, and at the Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, 710 Westwood Plaza, Los Angeles, California 90095-1769, USA.

e-mail: dhg@mednet.ucla.edu

1. Peltonen, L., Perola, M., Naukkarinen, J. & Palotie, A. *Hum. Mol. Genet.* **15**, R67–R74 (2006).
2. Jamain, S. *et al. Nature Genet.* **34**, 27–29 (2003).
3. Sebat, J. *et al. Science* **316**, 445–449 (2007).
4. Jacquemont, M. L. *et al. J. Med. Genet.* **43**, 843–849 (2006).
5. Strauss, K. A. *et al. N. Engl. J. Med.* **354**, 1370–1377 (2006).
6. Abrahams, B. S. & Geschwind, D. H. *Nature Rev. Genet.* **9**, 341–355 (2008).
7. Morrow, E. M. *et al. Science* **321**, 218–223 (2008).
8. Lander, E. S. & Botstein, D. *Science* **236**, 1567–1570 (1987).
9. Marshall, C. R. *et al. Am. J. Hum. Genet.* **82**, 477–488 (2008).
10. Alarcón, M. *et al. Am. J. Hum. Genet.* **82**, 150–159 (2008).
11. Flavell, S. W. & Greenberg, M. E. *Annu. Rev. Neurosci.* **31**, 563–590 (2008).
12. Geschwind, D. H. & Levitt, P. *Curr. Opin. Neurobiol.* **17**, 103–111 (2007).

## NEUROSCIENCE

# State-sanctioned synchrony

Scott J. Cruikshank and Barry W. Connors

**A sleepy brain pays little attention to its surroundings, and its neurons are lulled by a common oscillation. As the brain swiftly rouses from this sluggish state, its neurons function more independently.**

You know the experience. One moment you are drowsy and unfocused, awake but oblivious to your environment. Suddenly, something snaps you into awareness. Whether the trigger is external, like the sight or sound of someone entering the room, or internal, such as the thought of a looming deadline, you become immediately alert. Your ability to deal with most tasks, not just the stimulus that aroused you, improves, suggesting a change in your overall 'brain state'.

Understanding the neural mechanisms of brain states has been a goal of neuroscientists for decades<sup>1–3</sup>. One long-lived suggestion is that brain states relate to the degree to which neuronal activity in the cerebral cortex is synchronized<sup>4</sup>. On page 881 of this issue, Poulet and Petersen<sup>5</sup> directly address this issue, showing, in awake mice, that the electrical activity

of adjacent neurons becomes less synchronous when the brain enters an active state.

The long whiskers on a rodent's snout are important sensory organs. When mice are on the go, they sweep their whiskers through the local environment to feel and identify objects, and to navigate in darkness<sup>6</sup>. Poulet and Petersen<sup>5</sup> studied brain activity in mice during periods of active 'whisking', when the animals rhythmically waved their whiskers back and forth, and compared this with periods of 'quiet wakefulness', when the mice (and their whiskers) were almost motionless (Fig. 1a). The authors managed the technical feat of making simultaneous intracellular recordings from pairs of cortical neurons in awake, unanaesthetized animals. Intracellular recordings allow measurement not only of action potentials (the 'firing' of neurons, signalling their output), but

compared the sequence of *NHE9* in several hundred autistic subjects from non-consanguineous families and controls, and show that single base-pair mutations in the gene are associated with autism. Their finding therefore further supports a role for *NHE9* in increasing the susceptibility to this disorder.

The symptoms of autism include communication and social deficits, and restricted and repetitive behaviour. Clinical features can vary widely between individuals, and may include epilepsy or developmental regression after a period of apparently normal development. To simplify their analysis of *NHE9*, Morrow and colleagues focused on a subset of individuals who suffered from autism-associated epilepsy. It is thought that a subset of individuals who also have a particular additional disease feature (phenotype) in common might form a more genetically similar group. Unlike most of the autistic subjects without epilepsy, these individuals showed a more than fivefold increase in potentially harmful base-pair changes in the *NHE9* gene sequence, including a stop codon. Therefore, *NHE9* may be added to the list of genes that are associated with certain subsets of autism<sup>6</sup>.

Morrow and co-workers' observations<sup>7</sup> raise some intriguing questions. Five of the six homozygous deletions identified in the HMCA subjects lie in regions outside the coding sequence of the implicated genes, and so are unlikely to disrupt the encoded protein sequence. Several of the deletions might disrupt presumptive regulatory regions, whereas another 320-kilobase deletion closest to the *PCDH10* gene, which encodes the PCD10 (protocadherin 10) protein, does not lie in a coding region or in a sequence of known function. The effect of these deletions on neighbouring genes, and any causal role in autism, therefore require further investigation. Also, the causal nature of the *NHE9* mutations, which result in potentially deleterious changes at the protein level, must be established, because similar changes are observed in control subjects, albeit less frequently.

The difficulty in identifying mutations that are definitely pathogenic is an emerging theme in common neuropsychiatric disorders. Morrow and colleagues' study demonstrates that studying consanguineous families is a fruitful path to identifying autism genes, but also that such analysis is more challenging than might initially have been expected. This is partly due to the underlying genetic heterogeneity of autism, a facet underscored by this study. The authors find that no single recessive mutation is associated with autism in more than one family, even in consanguineous families from the same geographic region.

Luckily, advances in DNA sequencing technology mean that even large genomic regions can be efficiently sequenced, and work to identify recessive mutations in single base pairs within the genomes of subjects from the HMCA group might already be under way.

Sequencing shared regions in these families could be even more informative because many mutations in these regions are probably small — perhaps single base-pair mutations — and thus might more directly implicate specific genes.

Convergent evidence from other studies, including functional and gene-expression data, will be crucial in assessing the relevance of candidate genes to autism<sup>10</sup>. Morrow *et al.* embrace this notion, as their putative autism susceptibility genes represent diverse functional classes, which at face value do not suggest common molecular mechanisms. To provide a potential functional context, they show that three of the genes — either near (*PCDH10*) or overlapping (*c3orf58* and *NHE9*) the deletions — are among a significant subset that is associated with a form of neuronal activity thought to be involved in learning and memory<sup>11</sup>.

The authors suggest that investigating such genes might provide a way of unifying seemingly diverse molecular findings in autism. Given the nearly two dozen genes that are now associated with the disorder<sup>6</sup>, this should be a readily testable hypothesis. The way in which functional data relate to the core deficits of autism is not known, because many aspects of learning and memory are preserved in autistic children. The idea that autism might represent a general deficit in neuronal plasticity caused by changes in gene expression needs to be reconciled with the specificity of cognitive

dysfunction observed in the condition<sup>12</sup>, as well as the preservation of general intellectual ability in many individuals. This aspect highlights a central challenge of research in this area: connecting basic molecular pathways to the properties of complex neural circuits and cognition in humans. Spanning these diverse levels of function will require detailed analysis of the correlation between the genetic composition of individuals and their associated phenotype across the spectrum of autism disorders.

Daniel H. Geschwind is in the Program in Neurogenetics, Department of Neurology, and at the Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, 710 Westwood Plaza, Los Angeles, California 90095-1769, USA.

e-mail: dhg@mednet.ucla.edu

1. Peltonen, L., Perola, M., Naukkarinen, J. & Palotie, A. *Hum. Mol. Genet.* **15**, R67–R74 (2006).
2. Jamain, S. *et al. Nature Genet.* **34**, 27–29 (2003).
3. Sebat, J. *et al. Science* **316**, 445–449 (2007).
4. Jacquemont, M. L. *et al. J. Med. Genet.* **43**, 843–849 (2006).
5. Strauss, K. A. *et al. N. Engl. J. Med.* **354**, 1370–1377 (2006).
6. Abrahams, B. S. & Geschwind, D. H. *Nature Rev. Genet.* **9**, 341–355 (2008).
7. Morrow, E. M. *et al. Science* **321**, 218–223 (2008).
8. Lander, E. S. & Botstein, D. *Science* **236**, 1567–1570 (1987).
9. Marshall, C. R. *et al. Am. J. Hum. Genet.* **82**, 477–488 (2008).
10. Alarcón, M. *et al. Am. J. Hum. Genet.* **82**, 150–159 (2008).
11. Flavell, S. W. & Greenberg, M. E. *Annu. Rev. Neurosci.* **31**, 563–590 (2008).
12. Geschwind, D. H. & Levitt, P. *Curr. Opin. Neurobiol.* **17**, 103–111 (2007).

## NEUROSCIENCE

# State-sanctioned synchrony

Scott J. Cruikshank and Barry W. Connors

**A sleepy brain pays little attention to its surroundings, and its neurons are lulled by a common oscillation. As the brain swiftly rouses from this sluggish state, its neurons function more independently.**

You know the experience. One moment you are drowsy and unfocused, awake but oblivious to your environment. Suddenly, something snaps you into awareness. Whether the trigger is external, like the sight or sound of someone entering the room, or internal, such as the thought of a looming deadline, you become immediately alert. Your ability to deal with most tasks, not just the stimulus that aroused you, improves, suggesting a change in your overall 'brain state'.

Understanding the neural mechanisms of brain states has been a goal of neuroscientists for decades<sup>1–3</sup>. One long-lived suggestion is that brain states relate to the degree to which neuronal activity in the cerebral cortex is synchronized<sup>4</sup>. On page 881 of this issue, Poulet and Petersen<sup>5</sup> directly address this issue, showing, in awake mice, that the electrical activity

of adjacent neurons becomes less synchronous when the brain enters an active state.

The long whiskers on a rodent's snout are important sensory organs. When mice are on the go, they sweep their whiskers through the local environment to feel and identify objects, and to navigate in darkness<sup>6</sup>. Poulet and Petersen<sup>5</sup> studied brain activity in mice during periods of active 'whisking', when the animals rhythmically waved their whiskers back and forth, and compared this with periods of 'quiet wakefulness', when the mice (and their whiskers) were almost motionless (Fig. 1a). The authors managed the technical feat of making simultaneous intracellular recordings from pairs of cortical neurons in awake, unanaesthetized animals. Intracellular recordings allow measurement not only of action potentials (the 'firing' of neurons, signalling their output), but

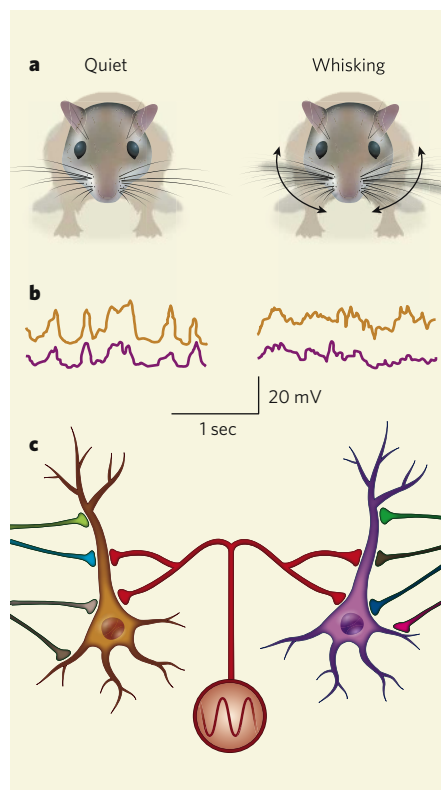
also of smaller electrical fluctuations, which mainly reflect inputs from other nerve cells. Poulet and Petersen<sup>5</sup> found that these sub-threshold fluctuations are highly correlated among cortical cells during quiet wakefulness and relatively desynchronized during whisking (Fig. 1b).

A correspondence between quiet brain states and strong cortical synchrony was the mainstream view of neuroscientists from the earliest days of electroencephalography (EEG, a method for recording extracellular voltages from the surface of the skull) until about 20 years ago. EEG pioneers referred to the large slow-waves associated with restful behaviour as 'synchronized'. Signals from individual neurons are tiny, so it was believed that synchronous activity of many neurons would be needed to give large measurable signals at EEG electrodes located outside the skull<sup>4,7</sup>. More recently, simultaneous EEG and intracellular recordings within the cortex during quiescent states (anaesthesia, sleep and quiet wakefulness) have demonstrated that slow oscillations of intracellular membrane potentials correlate well with EEG oscillations. Moreover, when brain states become active, either spontaneously or after stimulation, slow EEG waves and membrane-potential fluctuations diminish simultaneously<sup>8–10</sup>.

Because the EEG and the membrane potentials of most cortical neurons are strongly correlated, it seems very likely that slow membrane-potential oscillations are synchronized among cortical cells during quiescent states. In fact, synchrony among pairs of neurons has already been directly demonstrated in anaesthetized animals<sup>11,12</sup>. However, anaesthetics have powerful and poorly understood effects on neural circuits, and the cellular properties of the normally behaving brain have been particularly difficult to measure. This is why the work of Poulet and Petersen, measuring membrane potentials from pairs of cortical neurons in awake animals, is so important.

The decrease in membrane-potential correlation, observed when mice switched from quiet wakefulness to active whisking, was mainly due to decreased synchrony of low-frequency oscillations. During whisking, there was no obvious increase in relatively high-frequency (30–100 hertz) oscillations known as 'gamma' rhythms, or correlations of the same. This seems at odds with several previous studies of cortical activation (including one from the Petersen group<sup>8</sup>). In fact, since the late 1980s, an extensive literature has reported that activated brain states are accompanied by enhancement of gamma rhythms and their synchrony<sup>13–16</sup>.

The absence of enhanced gamma rhythms in the current study may reflect the ambiguous nature of the whisking state in these experiments. EEG measurements suggest that whisking periods might correspond to relatively active brain states. For instance, EEG fluctuations during quiet wakefulness have relatively



**Figure 1 | Brain state and cortical synchrony.** **a**, Quiet wakeful states of the mouse brain lack whisker movements; active brain states are indicated by whisking. **b**, Simultaneous recording from adjacent cortical neurons (pictured in **c**) show synchronous oscillations in membrane potential during quiet states. This synchrony breaks down during active (whisking) states. **c**, The synchronous, low-frequency cortical activity is driven by a neural 'pacemaker' (circle). The strength of this pacemaker is regulated by brain state (stronger during quiet wakefulness, weaker during whisking). Specific, information-rich pathways (various colours) generate relatively high-frequency, asynchronous synaptic inputs to neurons that dominate during active brain states.

high amplitudes and low frequencies (features classically associated with passive brain states), whereas the fluctuations during whisking tend to be smaller and more irregular<sup>8</sup> (features of active states). However, the heads of the mice had to be fixed in position for the duration of Poulet and Petersen's experiments<sup>5</sup>. No stimuli were provided, and the whisking apparently began and ended spontaneously, making the behavioural significance of this 'active' state uncertain<sup>6</sup>.

Poulet and Petersen<sup>5</sup> also show that eliminating sensory feedback from the whisker follicles, by severing the infraorbital nerve, affects neither state transitions themselves nor state-dependent changes in synchrony, even though all recordings are from the area of the cortex that processes this type of sensory information. The synaptic drives that mediate low-frequency synchrony are thus triggered centrally rather than by the direct action of the

sensory input on the cortex (Fig. 1c), consistent with much earlier studies<sup>2,17</sup>. However, Poulet and Petersen did not report whether the activation is cortex-wide (for example, including visual, auditory, prefrontal and limbic areas) or limited to whisker-processing regions. Some studies have shown that specific movements can be correlated with the localized activation of cortical areas associated with these motions<sup>18</sup>. Something similar may be happening here because the active state was defined by whisker movement itself. There may well be important mechanistic differences between the sort of activation produced by general 'arousal' systems of the brain<sup>2,3,9,17</sup>, and activation mediated by specific processes such as sensory-motor coordination.

This raises the issue of mechanisms more generally. Which neuromodulatory systems, if any, are involved in the state transitions? Where do the signals driving membrane-potential synchrony originate, and to what extent are they mediated by state-dependent excitatory or inhibitory processes<sup>19</sup>? And how do fluctuations in synchrony affect sensory processing<sup>6</sup>?

The relationship between brain state and cortical synchrony has been investigated for at least 80 years. This week's study by Poulet and Petersen<sup>5</sup> used direct measurements in a field that had been dominated by oblique techniques. This may, to paraphrase Winston Churchill, mark 'the end of the beginning' of an epic scientific problem.

Scott J. Cruikshank and Barry W. Connors are in the Department of Neuroscience, Brown University, Providence, Rhode Island 02912, USA. e-mails: scott\_cruikshank@brown.edu; barry\_connors@brown.edu

- Berger, H. *Arch. Psychiatr. Nervenkr.* **87**, 527–570 (1929).
- Moruzzi, G. & Magoun, H. W. *Electroencephalogr. Clin. Neurophysiol.* **1**, 455–473 (1949).
- Jones, B. E. *Ann. NY Acad. Sci.* **1129**, 26–34 (2008).
- Jasper, H. H. *Psychol. Bull.* **34**, 411–481 (1937).
- Poulet, J. F. A. & Petersen, C. C. H. *Nature* **454**, 881–885 (2008).
- Kleinfeld, D., Ahissar, E. & Diamond, M. E. *Curr. Opin. Neurobiol.* **16**, 435–444 (2006).
- Adrian, E. D. *Proc. R. Soc. Med.* **29**, 197–200 (1936).
- Crochet, S. & Petersen, C. C. H. *Nature Neurosci.* **9**, 608–610 (2006).
- Metherate, R., Cox, C. L. & Ashe, J. H. *J. Neurosci.* **12**, 4701–4711 (1992).
- Rudolph, M., Pospischil, M., Timofeev, I. & Destexhe, A. *J. Neurosci.* **27**, 5280–5290 (2007).
- Lampl, I., Reichova, I. & Ferster, D. *Neuron* **22**, 361–374 (1999).
- Petersen, C. C. H., Hahn, T. T. G., Mehta, M., Grinvald, A. & Sakmann, B. *Proc. Natl Acad. Sci. USA* **100**, 13638–13643 (2003).
- Gray, C. M., König, P., Engel, A. K. & Singer, W. *Nature* **338**, 334–337 (1989).
- Munk, M. H. J., Roelfsema, P. R., König, P., Engel, A. K. & Singer, W. *Science* **272**, 271–274 (1996).
- Steriade, M., Amzica, F. & Contreras, D. *J. Neurosci.* **16**, 392–417 (1996).
- Fries, P., Nikolić, D. & Singer, W. *Trends Neurosci.* **30**, 309–316 (2007).
- Lindsley, D. B., Schreiner, L. H., Knowles, W. B. & Magoun, H. W. *Electroencephalogr. Clin. Neurophysiol.* **2**, 483–498 (1950).
- Rougeul-Buser, A., Bouyer, J. J. & Buser, P. *Acta Neurobiol. Exp.* **35**, 805–819 (1975).
- Long, M. A., Cruikshank, S. J., Jutras, M. J. & Connors, B. W. *J. Neurosci.* **25**, 7309–7316 (2005).

# Genomics of cellulosic biofuels

Edward M. Rubin<sup>1,2</sup>

**The development of alternatives to fossil fuels as an energy source is an urgent global priority. Cellulosic biomass has the potential to contribute to meeting the demand for liquid fuel, but land-use requirements and process inefficiencies represent hurdles for large-scale deployment of biomass-to-biofuel technologies. Genomic information gathered from across the biosphere, including potential energy crops and microorganisms able to break down biomass, will be vital for improving the prospects of significant cellulosic biofuel production.**

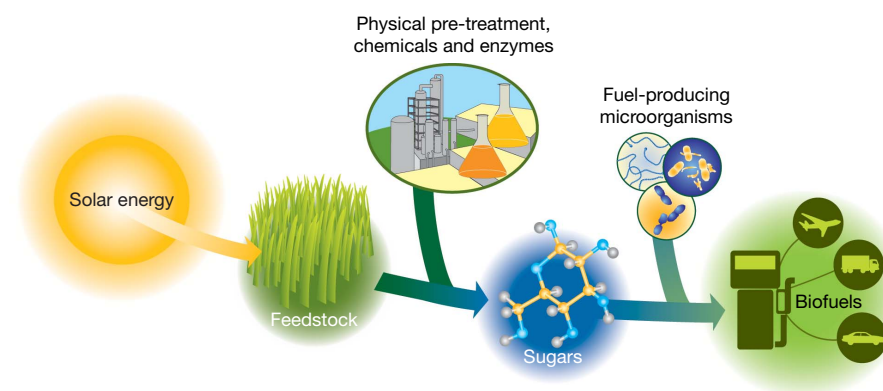
The capture of solar energy through photosynthesis is a process that enables the storage of energy in the form of cell wall polymers (that is, cellulose, hemicellulose and lignin). The energy stored in these polymers can be accessed in a variety of ways, ranging from simple burning to complex bioconversion processes. The high energy content and portability of biologically derived fuels, and their significant compatibility with existing petroleum-based transportation infrastructure, helps to explain their attractiveness as a fuel source. Despite the increasing use of biofuels such as biodiesel and sugar- or starch-based ethanol, evidence suggests that transportation fuels based on lignocellulosic biomass represent the most scalable alternative fuel source<sup>1</sup>. Lignocellulosic biomass in the form of plant materials (for example, grasses, wood and crop residues) offers the possibility of a renewable, geographically distributed and relatively greenhouse-gas-favourable source of sugars that can be converted to ethanol and other liquid fuels. Calculations of the productivity of lignocellulosic feedstocks, in part based on their ability to grow on marginal agricultural land, indicates that they can probably have a large impact on transportation needs without significantly compromising the land needed for food crop production<sup>2</sup>.

Lignocellulosic biofuel production involves collection of biomass, deconstruction of cell wall polymers into component sugars (pretreatment and saccharification), and conversion of the sugars to biofuels (fermentation) (Fig. 1). Partially because of the historically low demand for biologically based transportation fuels, each step in this process is in the early stages of optimization for efficiency and throughput. The crops from which biomass is currently derived have not been domesticated for this particular purpose and the present methods for saccharification and fermentation are inefficient and

expensive. However, the recent and pressing desire to develop alternatives to fossil fuels has made the rapid improvement of biofuel production a high priority, in which biologically derived energy ('bioenergy')-relevant genomic insights and resources will have an important role (Table 1).

## Biomass

From the perspective of transportation fuels, plants can be viewed as solar energy collectors and thermochemical energy storage systems. It is the storage of energy in a form that can later be accessed via thermochemical or enzymatic conversion that distinguishes biomass from other renewable energy sources. Cellulosic biomass, sometimes referred to as lignocellulosic biomass, is an abundant renewable resource that can be used for the production of alternative transportation fuels<sup>3</sup>. The three main components of lignocellulose are cellulose, hemicellulose and lignin (Fig. 2), with the relative proportions of the three dependent on the material source<sup>4</sup>. Cellulose, the main structural component of plant cell walls, is a long chain of glucose molecules, linked to one another primarily by glycosidic bonds<sup>5</sup>. Hemicellulose, the second most abundant constituent of lignocellulosic biomass, is not a chemically well defined compound but rather a family of polysaccharides, composed of different 5- and 6-carbon monosaccharide units, that links cellulose fibres into microfibrils and cross-links with lignin, creating a complex network of bonds that provide structural strength<sup>5</sup>. Finally lignin, a three-dimensional polymer of phenylpropanoid units, can be considered as the cellular glue providing the plant tissue and the individual fibres with compressive strength and the cell wall with stiffness<sup>6</sup>, in addition to providing resistance to insects and pathogens.



**Figure 1 | Biology of bioconversion of solar energy into biofuels.** Solar energy is collected by plants via photosynthesis and stored as lignocellulose. Decomposition of the cellulosic material into simple 5- and 6-carbon sugars is achieved by physical and chemical pre-treatment, followed by exposure to enzymes from biomass-degrading organisms. The simple sugars can be subsequently converted into fuels by microorganisms.

<sup>1</sup>DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598, USA. <sup>2</sup>Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA.

**Table 1 | Bioenergy genomes**

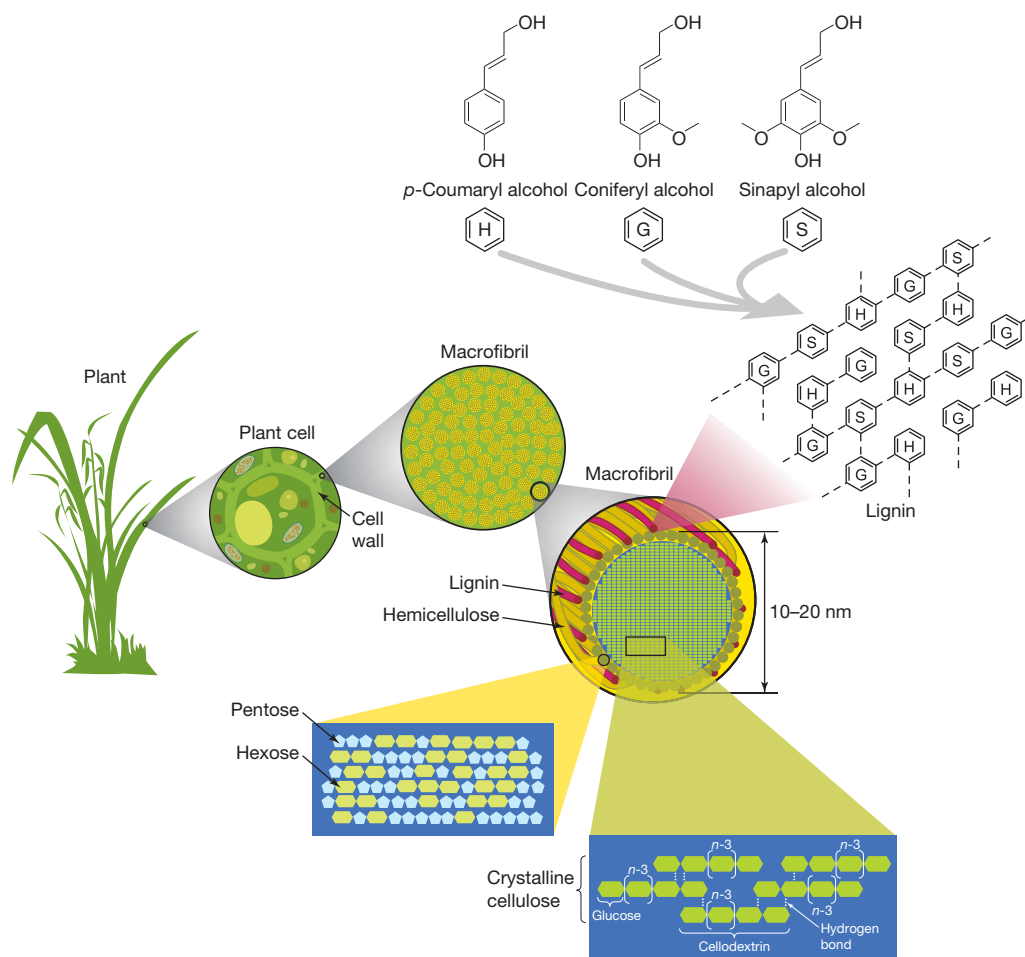
Organism	Genome size (megabases)	Status	Reference
<b>Feedstocks and feedstock models</b>			
<i>Populus trichocarpa</i> (poplar)	480	Complete	Ref. 9
<i>Chlamydomonas reinhardtii</i>	120	Complete	Ref. 34
<i>Glycine max</i> (soya bean)	1,200	Draft	–
<i>Manihot esculenta</i> (cassava)	770	In progress	–
<i>Sorghum bicolor</i>	760	In progress	–
<i>Eucalyptus globulus</i>	600	In progress	–
<i>Brachypodium distachyon</i>	355	In progress	–
<i>Zea mays</i> (maize)	2,500	In progress	–
<i>Elaeis guineensis</i> (oil palm)	~3,400	In progress	<a href="http://www.checkbiotech.org/green_News_Biofuels.aspx?infol=15100">http://www.checkbiotech.org/green_News_Biofuels.aspx?infol=15100</a>
<i>Panicum virgatum</i> (switchgrass)	~5,600	In progress	–
<i>Setaria italica</i> (foxtail millet)	~515	In progress	–
<b>Biomass degraders</b>			
<i>Acidothermus cellulolyticus</i> 11B	2.4	Complete	–
<i>Bacillus pumilus</i> SAFR-032	3.7	Complete	Ref. 35
<i>Caldicellulosiruptor saccharolyticus</i> DSM 8903	3.0	Complete	–
<i>Clostridium phytofermentans</i> ISDg	4.8	Complete	–
<i>Clostridium thermocellum</i> ATCC 27405	3.8	Complete	–
<i>Cytophaga hutchinsonii</i> ATCC 33406	4.4	Complete	–
<i>Flavobacterium johnsoniae</i> UW101	6.1	Complete	–
<i>Rubrobacter xylanophilus</i> DSM9941	3.2	Complete	–
<i>Saccharophagus degradans</i>	5.1	Complete	Ref. 36
<i>Thermobifida fusca</i> strain YX	3.6	Complete	Ref. 37
<i>Clostridium cellulolyticum</i> H10	4.0	Draft	–
<i>Elusimicrobium minutum</i> Pei191	1.6	Draft	–
<i>Nectria haematococca/Fusarium solani</i>	51	Draft	–
<i>Phanerochaete chrysosporium</i>	35.1	Draft	–
<i>Postia placenta</i>	33	Draft	–
<i>Sagittula stellata</i> E-37	5.3	Draft	–
<i>Trichoderma reesei/Hypocrea jecorina</i>	33	Draft	–
<i>Cellulomonas flavigena</i> DSM 20109	~4.0	In progress	–
<i>Cellvibrio japonicus</i> Ueda107	~6.0	In progress	–
<i>Fibrobacter succinogenes</i> subsp. <i>succinogenes</i> S85	~3.8	In progress	–
<i>Ruminococcus albus</i>	4.0	In progress	–
<i>Teredinibacter turnerae</i> T7902	~2	In progress	–
Termite hindgut community	NA	Complete	Ref. 23
Poplar biomass degrading community	NA	In progress	<a href="http://www.jgi.doe.gov/sequencing/lspssseqplans2007.html">http://www.jgi.doe.gov/sequencing/lspssseqplans2007.html</a>
Asian longhorned beetle ( <i>Anoplophora glabripennis</i> ) gut community	NA	In progress	<a href="http://www.jgi.doe.gov/sequencing/DOEmicrobes2007.html">http://www.jgi.doe.gov/sequencing/DOEmicrobes2007.html</a>
Bovine rumen community transcriptome	NA	In progress	<a href="http://www.energybiosciencesinstitute.org/index.php?option=com_content&amp;task=view&amp;id=159&amp;Itemid=20">http://www.energybiosciencesinstitute.org/index.php?option=com_content&amp;task=view&amp;id=159&amp;Itemid=20</a>
<b>Fuel producers</b>			
<i>Clostridium acetobutylicum</i> ATCC 824	4.0	Complete	Ref. 38
<i>Clostridium beijerinckii</i> NCIMB 8052	6.0	Complete	–
<i>Pichia stipitis</i>	15.4	Complete	Ref. 27
<i>Thermoanaerobacter tengcongensis</i> MB4	2.7	Complete	Ref. 39
<i>Zymomonas mobilis</i> subsp. <i>mobilis</i> ZM4	2.1	Complete	Ref. 40
<i>Bacillus coagulans</i> 36D1	2.9	Draft	–
<i>Thermoanaerobacter pseudethanolicus</i> 39E	2.4	Draft	–
<i>Clostridium ljungdahlii</i>	~4.0	In progress	–

Bioenergy-relevant organisms for which large-scale genome projects have been completed or are under way are listed. Information on genome projects without references can be found at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeproj>.

As we can retrospectively view the features that made certain wild plants desirable for domestication thousands of years ago to become today's food crops, we are now prospectively defining criteria to choose plants with potential to serve as dedicated bioenergy crops in the future. These include cell wall composition, growth rate, suitability for growth in different geographical regions, and resource-use efficiencies. With these features in mind, a list of potential bioenergy crops is being developed and targeted for different growing conditions<sup>7</sup>. Most plants assimilate their CO<sub>2</sub> first into a C<sub>3</sub> compound, whereas a smaller subset use a C<sub>4</sub> compound. Plants using C<sub>4</sub> photosynthesis tend to be among the most productive, having higher maximum efficiencies of light, nitrogen and water use in assimilating carbon. The C<sub>4</sub> group of potential energy crops includes various perennial grasses such as switchgrass and *Miscanthus*. These grasses have the advantages of not requiring replanting after a yearly harvest, rapid growth, high biomass density per unit area, and low nutrient and water needs, enabling growth on marginal agricultural land. Disadvantages are that C<sub>4</sub> plants are rare in cold climates and unable to grow at temperatures below 10 °C. In these environments, trees, which exclusively depend on C<sub>3</sub> photosynthesis, provide the only

candidate species. The C<sub>3</sub> group of potential energy crops includes trees, such as poplar and eucalyptus, which have relatively rapid growth potential in difficult-to-plough environments. It is highly likely that multiple different energy feedstocks will be deployed depending on latitude, geography, water availability and landowner acceptance.

Until recently, minimal effort has been directed towards optimizing potential energy crops for the generation of transportation fuels. This is in stark contrast to the agronomic development of food crops, which have been domesticated for thousands of years to maximize productivity. Teosinte, the wild precursor to modern maize, was first recognized by Native Americans more than 5,000 years ago as a potential food crop. The domestication of teosinte resulted in its conversion from a wild plant, the characteristics of which had been orchestrated by natural selection maximizing survival and reproduction, into a plant whose morphology and physiology had been extensively altered by artificial selection to increase its nutritional yield and ease of harvest<sup>8</sup>. More recently, selective breeding as well as agronomic advances have resulted in improvement over several orders of magnitude in the nutritional value per acre of modern maize



**Figure 2 | Structure of lignocellulose.** The main component of lignocellulose is cellulose, a  $\beta(1-4)$ -linked chain of glucose molecules. Hydrogen bonds between different layers of the polysaccharides contribute to the resistance of crystalline cellulose to degradation. Hemicellulose, the second most abundant component of lignocellulose, is composed of various 5- and 6-carbon sugars such as arabinose, galactose, glucose, mannose and xylose. Lignin is composed of three major phenolic components, namely

*p*-coumaryl alcohol (H), coniferyl alcohol (G) and sinapyl alcohol (S). Lignin is synthesized by polymerization of these components and their ratio within the polymer varies between different plants, wood tissues and cell wall layers. Cellulose, hemicellulose and lignin form structures called microfibrils, which are organized into macrofibrils that mediate structural stability in the plant cell wall.

compared to that of teosinte. Some of the most rapid increases have occurred in the past 40 years, both from advances in agronomic practices and, importantly, from the application of modern genetics. The optimization of bioenergy crops as feedstocks for transportation fuels is in its infancy, but already genomic information and resources are being developed that will be essential for accelerating their domestication. Many of the traits targeted for optimization in potential cellulosic energy crops are those that would improve growth on poor agricultural lands, to minimize competition with food crops over land use.

*Populus trichocarpa* (poplar), the first tree and potential bioenergy crop to have its genome sequenced (Table 1)<sup>9</sup>, illustrates some of the issues and potential of applying genomics to the challenge of optimizing energy crops. The traits for which the genetic underpinnings will be sought in the genomes of bioenergy-relevant plants, such as poplar, include those affecting growth rates, response to competition for light, branching habit, stem thickness and cell wall chemistry. Significant effort will go into maximizing biomass yield per unit land area, because this more than any other factor will minimize the impact on overall land use. One can imagine trees optimized to have short stature to increase light access and enable dense growth, large stem diameter, and reduced branch count to maximize energy density for transport and processing. Trees have evolved with highly rigid and stable cell walls due to heavy selective pressure for long life and an upright habit. Plants domesticated for energy production, with a

crop cycle time of only a few years, would have less need for a rigid cell wall than wild plants with lifetimes of a hundred years or more. Alterations in the ratios and structures of the various macromolecules forming the cell wall are a major target in energy crop domestication to facilitate post-harvest deconstruction at the cost of a less rigid plant.

Already, by comparing several of the presently available plant genomes (poplar<sup>9</sup>, rice<sup>10,11</sup>, *Arabidopsis*<sup>12</sup>; see Table 1) coupled with large-scale plant gene function and expression studies, a number of candidate genes for domestication traits have been identified<sup>13,14</sup>. These include many genes involved in cellulose and hemicellulose synthesis as well as those believed to influence various morphological growth characteristics such as height, branch number and stem thickness<sup>15</sup>. In addition to homology-based strategies, other genome-enabled strategies for identifying domestication candidate genes are being used. These include quantitative trait analysis of natural variation and genome-wide mutagenesis coupled with phenotypic screens for traits such as recalcitrance to sugar release, acid digestibility and general cell wall composition. The availability of high-throughput transgenesis in several plant systems<sup>16</sup> will facilitate functional studies to determine the *in vivo* activities of the large number of domestication candidate genes. Using these strategies, genes affecting features such as plant height, stem elongation and trunk radial growth, drought tolerance, and cell wall stability are but a few of the features that are likely to be identified as targets for domestication

in a fraction of the time required to carry out similar studies unaided by the plant genomes and genomic approaches<sup>17</sup>.

### Biomass degradation

The breakdown of biomass involves the release of long-chain polysaccharides, specifically cellulose and hemicellulose, and the subsequent hydrolysis of these polysaccharides into their component 5- and 6-carbon chain sugars. Early chemical processes for cellulose degradation depended heavily on acid treatments<sup>5,18</sup>, and even today industrial processes for biomass degradation involve heat and acidic conditions and tend to be expensive, slow and relatively inefficient<sup>19</sup>. Furthermore, some of these pre-treatments produce inhibitors (that is, furfural and 5-hydroxymethylfurfural) that decrease the overall yield of the fermentation process.

The human need for efficient breakdown of lignocellulosic biomass for biofuel production is relatively new; however, a variety of organisms have evolved to take advantage of this nutrient source, including the free-living organisms and symbiotic animal-microbe consortia invariably present in biomass-rich environments. Increasing our knowledge of the biochemical machinery used by these organisms for the breakdown of biomass offers new avenues for the development of biologically based processes that could potentially accomplish biomass conversion at an industrial scale. Just as plant genomes are providing us with new approaches for accelerating feedstock domestication, genomics of biomass-degrading microbes and microbial consortia offer a new means to enzyme discovery.

Microbial strategies for degrading lignocellulose are diverse, yet our current understanding of the enzymes involved in these processes is limited to a handful of model organisms such as the fungus *Trichoderma reesei* and the bacterium *Clostridium thermocellum*<sup>20</sup>. Recent genome sequencing of these two biomass degraders, as well as a host of other cellulolytic species (Table 1), has expanded our repertoire of known or potential cellulolytic enzymes. The knowledge of biomass degradation pathways is soon to be increased even more by a number of large-scale genomic surveys under way of isolated organisms and microbial communities known to degrade biomass (Table 1).

Considering the current dependence on acid and heat pre-treatment in the deconstruction of lignocellulose, it is clear that enzymes that are stable and active at low pH values and at high temperatures are of particular value. Thus, enzymes derived from thermophilic and acidophilic organisms known to degrade lignocellulose, such as *Caldicellulosiruptor saccharolyticus* and *Acidothermus cellulolyticus*, hold significant promise for industrial processes<sup>21</sup>. However, the number of sequenced thermoacidophilic organisms is low, partly because of the significant obstacles in cultivating these organisms in the laboratory.

Many novel enzymes and enzyme systems that have evolved to make use of cellulosic biomass are present in difficult-to-culture microbes<sup>22</sup>. This is particularly the case for the communities of microbes inhabiting the guts of lignocellulose-consuming insects and ruminants. Recently, in response to the challenge of studying difficult-to-culture organisms, metagenomic approaches have been developed to access the information encoded in their genomes (Table 1). This typically involves the high-throughput sequencing of DNA extracted directly from the mixtures of organisms present in an environment.

The metagenomic analysis of a wood-feeding higher termite<sup>23</sup> has recently revealed hundreds of genes encoding carbohydrate-active enzymes, many of which were subsequently shown to be biochemically active. From the perspective of an industrial application, the termite hindgut is too complex a model to mimic because it comprises hundreds of microbial species and lignocellulose-degrading enzymes. However, termites tend to feed on complex diets of several plant species, which translate to more complex hindgut communities and enzyme inventories. Thus, examination of this single community reveals enzymes capable of hydrolysing a broad assortment of chemical

bonds. The saccharification of a single uniform cellulosic feedstock, particularly one engineered to have a simplified cell wall, will presumably require fewer enzymes and less downstream optimization of gene regulation and enzyme activity in a production host.

The availability of a wide range of naturally occurring lignocellulose-degrading enzymes increases the chances of successful enzyme optimization for industrial processes. Optimization of the saccharification process is crucial because the cost of cellulases remains a key barrier to economical production of biofuels<sup>3</sup>. A more diverse set of candidate enzymes—identified through a combination of conventional cultured microbial genome studies coupled with environmental prospecting methods such as metagenomics—improves the likelihood of obtaining enzymes with activities and stability suited to a variety of industrial processes.

### Biofuel production

Although the final steps of cellulosic ethanol production will require much of the same infrastructure developed for the production of sugar- and starch-based ethanol, changes will need to be made to exploit the diversity of sugars generated from the breakdown of biomass. Whereas the conversion of starch-based feedstock results primarily in hexoses, the products from degradation of lignocellulosic biomass, composed in part of hemicellulose, include large amounts of the pentose sugars D-xylose and L-arabinose. In contrast to the hexose sugars, the pentose sugars cannot be fermented by wild-type *Saccharomyces cerevisiae*<sup>24</sup>. Another factor limiting ethanol yield is the toxicity of ethanol to the fermenting host. Most fermenting organisms such as *S. cerevisiae* cannot tolerate ethanol concentrations exceeding 25% (v/v)<sup>25</sup>, resulting in a product that must then be concentrated through distillation. Distillation represents an expensive and energy-intensive step in ethanol production.

These limitations in ethanol production have generated significant interest in developing new organisms able to exploit fully the breakdown products of lignocellulosic biomass as well as tolerate the products of fermentation<sup>26</sup>. *Pichia stipitis* represents one yeast species of relevance to biofuel research based on its natural ability to ferment the pentose sugar xylose. Its recently sequenced genome revealed insights into the metabolic pathways responsible for this process<sup>27</sup>, and investigators are already working on the optimization of this pathway in *P. stipitis* as well as on the construction of systems for the heterologous expression of *P. stipitis* genes. *Escherichia coli* has already been genetically engineered for the conversion of all hexose and pentose sugars present in hemicellulose polymers<sup>28</sup>. The resulting genetically engineered strain has an ethanol production rate similar to yeast<sup>29</sup>, and its ethanol tolerance has been increased by selection on enrichment media<sup>30</sup>. In the future, genomics and pathway engineering should considerably facilitate the development of a variety of organisms able to use the full repertoire of cellulosic and hemicellulosic sugars and tolerate high ethanol concentrations to optimize ethanol yields.

Although ethanol production facilities and distribution centres are rapidly increasing in number, ethanol as a transportation fuel has several disadvantages that encourage the pursuit of better suited alternative biofuels in the future. Disadvantages of ethanol as a biofuel include: (1) high solubility of ethanol in water, which necessitates an energy-intensive distillation step; (2) diminished energy per gallon content of ethanol compared to petroleum; (3) difficulties in the distribution of ethanol via today's pipeline infrastructure owing to its hygroscopic nature; and (4) incompatibilities with many current vehicles at higher blending volumes. Thus, work is under way to enable the synthesis of next-generation biofuel products, such as higher-chain alcohols, alkanes, and other molecules with structures and activities closer to that of petroleum and diesel. Butanol, for example, can be used in conventional engines either in pure form or mixed with petroleum, can be transported by existing pipelines, and has a higher energy content than petroleum<sup>31</sup>. A number of organisms, such as *Clostridium acetobutylicum*, produce butanol

from sugars via the ABE (acetone, butanol, ethanol) fermentation and can be manipulated to produce mostly butanol<sup>31</sup>. *E. coli* has recently been engineered to produce isobutanol and other alcohols via a non-fermentative pathway that may be more readily adapted to large-scale production, using heterologous expression of *Lactococcus lactis* and *Bacillus subtilis* genes<sup>32</sup>. The genomics-enabled construction of partial or fully artificial biological systems through 'synthetic biology' approaches will be key in developing efficient, inexpensive biofuel production systems<sup>33</sup>.

## Summary

This perspective has focused on biofuels derived from lignocellulosic feedstocks; however, another potential source of transportation fuels is biodiesel, derived from oil-producing plants and microbes. Predominant sources of biodiesel presently include soya bean, rapeseed and palm oils. The genomes of soya bean and oil palm are both currently being sequenced (Table 1), information from which will be called on to increase oil production and facilitate plant growth in a broader range of habitats. Oil-producing algae are also under active study as an alternative source for biodiesel production. One of the major advantages of algae is their growth in liquid, negating the issue of potential competition between food and energy crops for land use. *Chlamydomonas reinhardtii* is one of a series of algal species whose genomes are becoming available for analysis<sup>34</sup>. Considering the magnitude of the transportation fuel problem, biodiesel is likely to contribute at least in part to the solution, and genomic research on oil producers will aid in making biodiesel production cheaper and more efficient.

Placing a man on the Moon and sequencing the human genome represented large-scale technologically challenging projects. In both these instances, substantial research and development efforts resulted in technological advances unforeseen at the projects' initiation. The strategies that were pioneered in sequencing and interpreting the human genome for the improvement of human health are now poised to be an important contributing technology in the challenge to develop environmentally and socially acceptable alternatives to fossil fuels.

- Hill, J., Nelson, E., Tilman, D., Polasky, S. & Tiffany, D. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proc. Natl Acad. Sci. USA* **103**, 11206–11210 (2006).
- Tilman, D., Hill, J. & Lehman, C. Carbon-negative biofuels from low-input high-diversity grassland biomass. *Science* **314**, 1598–1600 (2006).  
A demonstration of why grassland perennials, such as switchgrass, are superior for biofuel production when compared to crops that presently serve as food crops, such as soya bean or maize.
- Himmel, M. E. et al. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315**, 804–807 (2007).
- Reddy, N. & Yang, Y. Biofibers from agricultural byproducts for industrial applications. *Trends Biotechnol.* **23**, 22–27 (2005).
- van Wyk, J. P. Biotechnology and the utilization of biowaste as a resource for bioproduct development. *Trends Biotechnol.* **19**, 172–177 (2001).  
A discussion of biowaste as a potential source of lignocellulose for biofuel production.
- Del Rio, J. C., Marques, G., Rencoret, J., Martinez, A. T. & Gutierrez, A. Occurrence of naturally acetylated lignin units. *J. Agric. Food Chem.* **55**, 5461–5468 (2007).
- Sanderson, K. US biofuels: a field in ferment. *Nature* **444**, 673–676 (2006).
- Doebley, J. F., Gaut, B. S. & Smith, B. D. The molecular genetics of crop domestication. *Cell* **127**, 1309–1321 (2006).
- Tuskan, G. A. et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604 (2006).
- Yu, J. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79–92 (2002).
- Goff, S. A. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* **296**, 92–100 (2002).
- The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
- Kalluri, U. C., Difazio, S. P., Brunner, A. M. & Tuskan, G. A. Genome-wide analysis of Aux/IAA and ARF gene families in *Populus trichocarpa*. *BMC Plant Biol.* **7**, 59 (2007).
- Busov, V. B., Brunner, A. M. & Strauss, S. H. Genes for control of plant stature and form. *New Phytol.* **177**, 589–607 (2008).
- Ragauskas, A. J. et al. The path forward for biofuels and biomaterials. *Science* **311**, 484–489 (2006).
- Filichkin, S. A. et al. Efficiency of gene silencing in *Arabidopsis*: direct inverted repeats vs. transitive RNAi vectors. *Plant Biotechnol. J.* **5**, 615–626 (2007).
- Dinus, R. J., Payne, P., Sewell, M. M., Chiang, V. L. & Tuskan, G. A. Genetic modification of short rotation poplar wood properties for energy and fiber production. *Crit. Rev. Plant Sci.* **20**, 51–69 (2001).
- LaForge, F. B. & Hudson, C. S. The preparation of several useful substances from corn cobs. *J. Ind. Eng. Chem.* **10**, 925–927 (1918).
- Mosier, N. et al. Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour. Technol.* **96**, 673–686 (2005).
- Gilbert, H. J. Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. *Mol. Microbiol.* **63**, 1568–1576 (2007).
- Viikari, L., Alapuranen, M., Puranen, T., Vehmaanpera, J. & Siika-Aho, M. Thermostable enzymes in lignocellulose hydrolysis. *Adv. Biochem. Eng. Biotechnol.* **108**, 121–145 (2007).
- Hugenholtz, P. Exploring prokaryotic diversity in the genomic era. *Genome Biol.* **3**, REVIEWS0003 (2002).
- Warnecke, F. et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**, 560–565 (2007).  
A metagenomic study of an invertebrate gut microbial community involved in lignocellulolytic degradation.
- van Maris, A. J. et al. Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: current status. *Antonie Van Leeuwenhoek* **90**, 391–418 (2006).
- Wang, M., Zhao, J., Yang, Z. & Du, Z. Electrochemical insights into the ethanol tolerance of *Saccharomyces cerevisiae*. *Bioelectrochemistry* **71**, 107–112 (2007).
- Georgieva, T. I., Mikkelsen, M. J. & Ahning, B. K. High ethanol tolerance of the thermophilic anaerobic ethanol producer *Thermoanaerobacter* BG1L1. *Central Eur. J. Biol.* **2**, 364–377 (2007).
- Jeffries, T. W. et al. Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nature Biotechnol.* **25**, 319–326 (2007).
- Ohta, K., Beall, D. S., Mejia, J. P., Shanmugam, K. T. & Ingram, L. O. Genetic improvement of *Escherichia coli* for ethanol production: chromosomal integration of *Zymomonas mobilis* genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. *Appl. Environ. Microbiol.* **57**, 893–900 (1991).  
A description of the genetic modification of *E. coli*, yielding a strain capable of fermenting pentose and hexose sugars—which are present in lignocellulose—into ethanol.
- Jarboe, L. R., Grabar, T. B., Yomano, L. P., Shanmugam, K. T. & Ingram, L. O. Development of ethanologenic bacteria. *Adv. Biochem. Eng. Biotechnol.* **108**, 237–261 (2007).
- Yomano, L. P., York, S. W. & Ingram, L. O. Isolation and characterization of ethanol-tolerant mutants of *Escherichia coli* K011 for fuel ethanol production. *J. Ind. Microbiol. Biotechnol.* **20**, 132–138 (1998).
- Durre, P. Biobutanol: an attractive biofuel. *Biotechnol. J.* **2**, 1525–1534 (2007).
- Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–89 (2008).
- Lartigue, C. et al. Genome transplantation in bacteria: changing one species to another. *Science* **317**, 632–638 (2007).
- Merchant, S. S. et al. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* **318**, 245–250 (2007).
- Gioia, J. et al. Paradoxical DNA repair and peroxide resistance gene conservation in *Bacillus pumilus* SAFR-032. *PLoS ONE* **2**, e928 (2007).
- Taylor, L. E. et al. Complete cellulase system in the marine bacterium *Saccharophagus degradans* strain 2-40T. *J. Bacteriol.* **188**, 3849–3861 (2006).
- Lykidis, A. et al. Genome sequence and analysis of the soil cellulolytic actinomycete *Thermobifida fusca* YX. *J. Bacteriol.* **189**, 2477–2486 (2007).
- Nolling, J. et al. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183**, 4823–4838 (2001).
- Bao, Q. et al. A complete sequence of the *T. tengcongensis* genome. *Genome Res* **12**, 689–700 (2002).
- Seo, J. S. et al. The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. *Nature Biotechnol.* **23**, 63–68 (2005).

**Acknowledgements** I would like to thank S. Tringe, M. Hess, J. Tuskan, P. Hugenholtz, J. Bristow, B. Simmons, S. Long, J. Fruchart-Najib and H. Blanch for their input to the manuscript. This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract number DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract number DE-AC52-07NA27344, and Los Alamos National Laboratory under contract number DE-AC02-06NA25396.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence should be addressed to E.M.R. ([emrubin@lbl.gov](mailto:emrubin@lbl.gov)).

## ARTICLES

# UCP2 mediates ghrelin's action on NPY/AgRP neurons by lowering free radicals

Zane B. Andrews<sup>1,2</sup>, Zhong-Wu Liu<sup>2,5</sup>, Nicholas Wallingford<sup>2</sup>, Derek M. Erion<sup>2</sup>, Erzsebet Borok<sup>2</sup>, Jeffery M. Friedman<sup>6</sup>, Matthias H. Tschöp<sup>7</sup>, Marya Shanabrough<sup>2</sup>, Gary Cline<sup>3</sup>, Gerald I. Shulman<sup>3</sup>, Anna Coppola<sup>2</sup>, Xiao-Bing Gao<sup>2</sup>, Tamas L. Horvath<sup>1,2,4</sup> & Sabrina Diano<sup>2,4</sup>

**The gut-derived hormone ghrelin exerts its effect on the brain by regulating neuronal activity. Ghrelin-induced feeding behaviour is controlled by arcuate nucleus neurons that co-express neuropeptide Y and agouti-related protein (NPY/AgRP neurons). However, the intracellular mechanisms triggered by ghrelin to alter NPY/AgRP neuronal activity are poorly understood. Here we show that ghrelin initiates robust changes in hypothalamic mitochondrial respiration in mice that are dependent on uncoupling protein 2 (UCP2). Activation of this mitochondrial mechanism is critical for ghrelin-induced mitochondrial proliferation and electric activation of NPY/AgRP neurons, for ghrelin-triggered synaptic plasticity of pro-opiomelanocortin-expressing neurons, and for ghrelin-induced food intake. The UCP2-dependent action of ghrelin on NPY/AgRP neurons is driven by a hypothalamic fatty acid oxidation pathway involving AMPK, CPT1 and free radicals that are scavenged by UCP2. These results reveal a signalling modality connecting mitochondria-mediated effects of G-protein-coupled receptors on neuronal function and associated behaviour.**

Ghrelin alters neuronal activity in various brain regions including the hypothalamus<sup>1</sup>, hippocampus<sup>2</sup> and midbrain<sup>3</sup>. Ghrelin activates NPY/AgRP neurons of the arcuate nucleus<sup>1,4</sup> to suppress pro-opiomelanocortin (POMC) neuronal activity while promoting food intake. However, the intracellular mechanisms that mediate NPY/AgRP neuronal activation in response to ghrelin remain unknown.

Uncoupling proteins are inner-membrane mitochondrial proteins that are important for bioenergetics of diverse tissues<sup>5–8</sup>. UCP2 is strongly expressed in the hypothalamus including NPY/AgRP neurons of the arcuate nucleus<sup>9,10</sup> where ghrelin receptors (growth hormone secretagogue receptor 1; GHSR) are also expressed<sup>11</sup>. Ghrelin enhances *Ucp2* mRNA expression in the pancreas<sup>12</sup>, liver<sup>13</sup> and white adipose tissue<sup>14</sup>, and hypothalamic *Ucp2* mRNA levels and protein activity are elevated during fasting<sup>9</sup>, when circulating ghrelin is increased<sup>15</sup>. Thus, we analysed whether UCP2 is involved in hypothalamic neuronal responses to ghrelin and associated food intake.

## Ghrelin induces mitochondrial respiration via UCP2

Immunohistochemistry revealed the presence of GHSR in greater than 90% of UCP2-expressing,  $\beta$ -galactosidase-positive arcuate nucleus neurons in *Ucp2*<sup>−/−</sup> (ref. 16) mice (Supplementary Fig. 1a–c).

To analyse whether ghrelin affects mitochondrial respiration, we studied isolated hypothalamic synaptosomes<sup>17</sup> in ghrelin- and saline-treated animals. Mitochondrial respiration (given as nmol of oxygen consumed per minute per mg protein) was significantly increased after adding pyruvate and malate (state 2; saline 32.07 ± 1.38 and ghrelin 44.83 ± 3.15,  $P < 0.05$ , wild-type mice). The subsequent addition of ADP (state 3) increased respiration in the ghrelin-treated mice relative to saline controls (saline 99.39 ± 3.51 and ghrelin 155.98 ± 5.36,  $P < 0.05$ , wild-type mice). Ghrelin increased basal uncoupling status after ATP synthase inhibition with oligomycin (saline 27.2 ± 2.2 and ghrelin 37.3 ± 1.6,  $P < 0.05$ , wild-type mice;

Fig. 1a) and free-fatty-acid (palmitate)-induced respiration (saline 150.0 ± 5.7 and ghrelin 170.1 ± 4.7,  $P < 0.05$ , wild-type mice; Fig. 1b). Finally, total respiratory capacity induced by carbonylcyanide-4-(trifluoromethoxy)-phenylhydrazone (FCCP) was also increased by ghrelin (saline 297.0 ± 7.7 and ghrelin 427.9 ± 19.2,  $P < 0.05$ , wild-type mice; Fig. 1c). These increases were dependent on UCP2, as no significant elevation was seen for the above respiration parameters in *Ucp2*<sup>−/−</sup> mice compared to their saline controls (Fig. 1a–c). Genipin, a compound reported to be a specific UCP2 inhibitor<sup>18</sup>, inhibited oxygen consumption in all states of mitochondrial respiration in all tissues studied from both wild-type and *Ucp2*<sup>−/−</sup> mice (Supplementary Table 1).

In assessing whether ghrelin's effect on UCP2 activation involves *Ucp2* gene expression, we found that ghrelin produced a robust increase in hypothalamic *Ucp2* mRNA in wild-type mice, but not in *Gshr*<sup>−/−</sup> mice (Supplementary Fig. 1d). Ghrelin also stimulated *Npy* gene expression in the arcuate nucleus in a UCP2- (Fig. 1e) and GHSR-dependent (Supplementary Fig. 1e) manner. Finally, ghrelin-induced AgRP expression also depended on UCP2 (Supplementary Fig. 1f). Ghrelin had no effect on POMC (Supplementary Fig. 1g).

## Ghrelin induces UCP2-mediated mitochondrial proliferation

During fasting the number of mitochondria in NPY/AgRP perikarya increases in a UCP2-dependent manner<sup>9</sup>. Ghrelin, which is elevated during fasting, also increased mitochondrial number in NPY neuronal perikarya in wild-type but not *Ucp2*<sup>−/−</sup> mice (Fig. 1f and Supplementary Fig. 1h, i). The number of mitochondria in POMC-GFP perikarya was unaffected (Supplementary Fig. 1j). Because we observed a ghrelin-induced, UCP2- and GHSR-dependent increase in nuclear respiratory factor 1 (*Nrf1*; Fig. 1g) expression, a transcription factor that stimulates mitochondrial biogenesis, we propose that the change in mitochondrial number in NPY/AgRP

<sup>1</sup>Section of Comparative Medicine and Departments of, <sup>2</sup>Obstetrics, Gynecology & Reproductive Sciences, <sup>3</sup>Internal Medicine, Howard Hughes Medical Institute and <sup>4</sup>Neurobiology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. <sup>5</sup>Department of Neurobiology, Yunnan Medical College, Shiyuan, Hubei 442000, China. <sup>6</sup>Laboratory of Molecular Genetics, Howard Hughes Medical Institute, Rockefeller University, New York, New York 10021, USA. <sup>7</sup>Department of Psychiatry, University of Cincinnati, Cincinnati, Ohio 45237, USA.

neurons may also involve *de novo* mitochondrial biogenesis. This is supported by the observation that peroxisome proliferator-activated receptor- $\gamma$  coactivator 1 $\alpha$  (PGC1 $\alpha$ ) enhances mitochondrial biogenesis and respiration in association with robust induction of UCP2 and NRF1 (ref. 8).

The above results on perikaryal mitochondria number and from counts of synaptosomal mitochondria from hypothalamic tissue of

wild-type or *Ucp2*<sup>-/-</sup> mice treated with ghrelin or saline (see Supplementary Fig. 2a–g), together with ghrelin's effect on state 3 respiration and total respiration capacity of synaptosomal mitochondria, suggest that elevated oxygen consumption by mitochondria in response to ghrelin arises in part from increased mitochondrial number.

### Ghrelin alters mitochondrial membrane potential via UCP2

In peripheral tissues<sup>19,20</sup> UCP2 alters mitochondrial membrane potential. Thus, we analysed the mitochondrial membrane potential in synaptosomes of ghrelin- or saline-treated wild-type and *Ucp2*<sup>-/-</sup> mice. Ghrelin-treated wild-type synaptosomes exhibited lower mitochondrial membrane potential after addition of pyruvate and malate (state 2:  $78.43 \pm 8.61\%$  of saline values;  $P < 0.05$ ) and oligomycin (state 4:  $75.13 \pm 4.87\%$  of saline values;  $P < 0.05$ ) compared to saline-treated wild-type synaptosomes (Fig. 1h). These differences were absent in *Ucp2*<sup>-/-</sup> mice (Supplementary Fig. 3a). These observations suggest that a component of the altered oxygen consumption induced by ghrelin may also arise from increased uncoupled respiration; however, further studies are needed to strengthen these findings and study the basal differences in mitochondrial membrane potential kinetics in wild-type and *Ucp2*<sup>-/-</sup> mice.

### Ghrelin-induced NPY/AgRP activation is dependent on UCP2

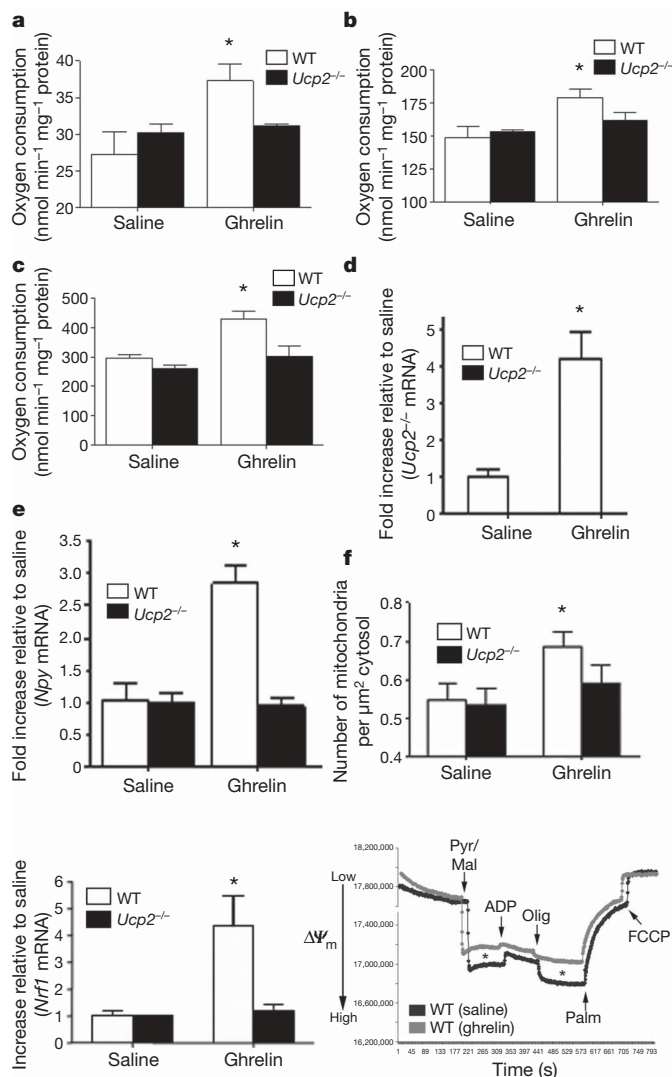
As a surrogate of neuronal activation, c-fos expression was examined in response to ghrelin in NPY neurons using mice expressing NPY-GFP on the *Ucp2*<sup>-/-</sup> background (NPY-GFP/*Ucp2*<sup>-/-</sup> mice). After ghrelin injection, total NPY/c-fos dual labelling was significantly attenuated in NPY-GFP/*Ucp2*<sup>-/-</sup> mice compared to NPY-GFP/wild-type mice (wild type  $602.33 \pm 98.50$ , *Ucp2*<sup>-/-</sup>  $298.40 \pm 64.59$ ,  $P < 0.05$ ; Fig. 2a and Supplementary Fig. 3b–d). Neither total NPY cell number nor arcuate nucleus volume was affected by genotype or ghrelin treatment (Supplementary Fig. 3e, f). No difference was observed in POMC-GFP/*Ucp2*<sup>-/-</sup> mice (Supplementary Fig. 3g–i).

Next, patch-clamp whole-cell electrophysiological recordings in slice preparations from NPY-GFP/wild type and NPY-GFP/*Ucp2*<sup>-/-</sup> mice showed that ghrelin increased action potential firing in all NPY neurons; however, it was severely attenuated in *Ucp2*<sup>-/-</sup> mice compared to wild-type animals (Fig. 2b and Supplementary Fig. 4a). This defect was specific to the action of ghrelin on NPY neurons, as glutamate produced normal responses in wild-type and *Ucp2*<sup>-/-</sup> mice (Supplementary Fig. 4b). Ghrelin had no effect on POMC-GFP neuronal action potential frequency. These results are concordant with anatomical observations showing that 94% of NPY and just 8% of POMC neurons express GHSRs<sup>11</sup>.

### Ghrelin-induced synaptic plasticity is mediated by UCP2

Ghrelin indirectly hyperpolarizes POMC neurons by activating inhibitory NPY/AgRP inputs<sup>1</sup>, and, consistent with this, ghrelin-induced recruitment of miniature inhibitory postsynaptic currents (mIPSCs) on POMC cells was significantly elevated in wild-type animals ( $153.8 \pm 22.4\%$ ,  $n = 5$ ), but not in *Ucp2*<sup>-/-</sup> mice ( $100.5 \pm 2.5\%$ ,  $n = 5$ ; Fig. 2c and Supplementary Fig. 4c).

Changes in mIPSCs are associated with synaptic plasticity<sup>21–23</sup>. Corresponding to the mIPSC measurements, ghrelin treatment increased symmetric, putatively inhibitory synapses on POMC neurons in wild-type (saline  $8.33 \pm 0.97$  and ghrelin  $16.55 \pm 1.68$ ) but not in *Ucp2*<sup>-/-</sup> mice (saline  $7.23 \pm 1.63$  and ghrelin  $9.12 \pm 1.63$ ; Fig. 2d). Conversely, ghrelin significantly reduced the number of asymmetric, putatively excitatory synapses on POMC neurons in wild-type mice (saline  $15.75 \pm 1.89$  and ghrelin  $8.66 \pm 1.05$ ) but not in *Ucp2*<sup>-/-</sup> mice (Supplementary Fig. 4d). Although the inhibitory input of POMC neurons arises partially from arcuate nucleus NPY/AgRP neurons, the origin of the stimulatory synapses needs further exploration.

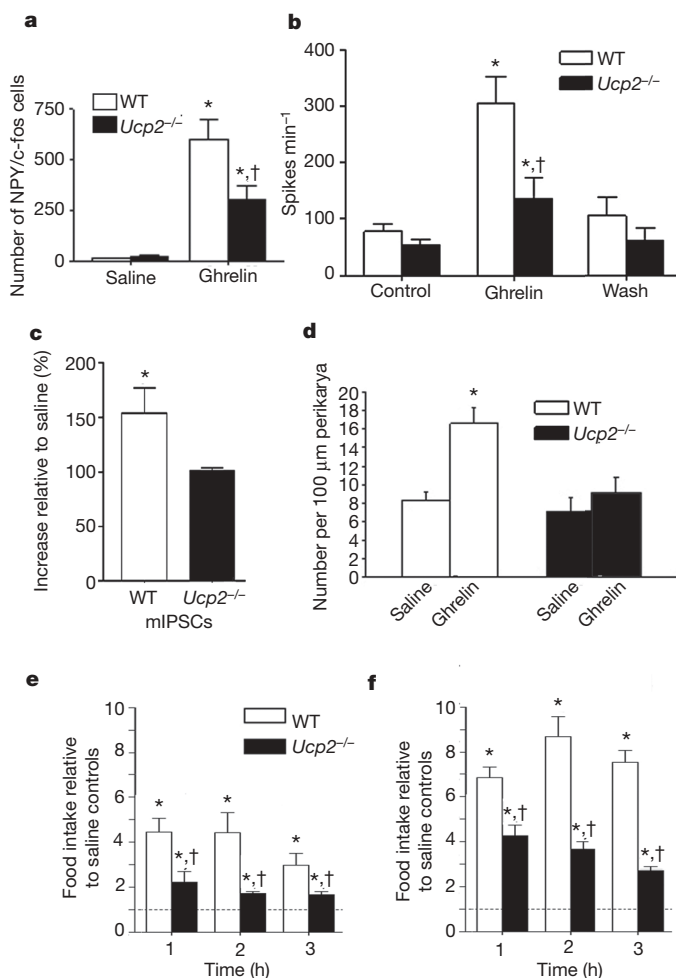


**Figure 1 | Mitochondrial and transcriptional effects of ghrelin are mediated by UCP2.** a–c, Ghrelin alters hypothalamic mitochondrial respiration.

a, Oligomycin inhibits ATP synthase and uncouples substrate oxidation from ATP phosphorylation. The resulting increase in uncoupled oxygen consumption in ghrelin-treated mice is due to UCP2, as *Ucp2*<sup>-/-</sup> mice exhibit no response to ghrelin. b, Ghrelin enhances respiration after palmitate addition in wild-type but not *Ucp2*<sup>-/-</sup> mice. c, An increase in respiration after FCCP shows that ghrelin enhances the maximal respiratory capacity in wild-type but not *Ucp2*<sup>-/-</sup> mice. d, Ghrelin increases (10 nmol) hypothalamic *Ucp2* mRNA expression as measured by real-time PCR ( $n = 6$ ). The data are expressed as a fold increase relative to saline levels. e, Arcuate *Npy* gene expression is increased in response to ghrelin in wild-type but not *Ucp2*<sup>-/-</sup> mice ( $n = 6$ ). f, Ghrelin promotes mitochondrial proliferation in a UCP2-dependent manner in arcuate NPY/AgRP cells. g, Ghrelin increases *Nrf1* mRNA expression in the hypothalamus of wild-type but not *Ucp2*<sup>-/-</sup> mice. h, Representative traces showing that ghrelin decreases the mitochondrial membrane potential after addition of energy substrates (state 2) and after addition of oligomycin (state 4). Asterisk indicates statistically significant differences ( $P < 0.05$ ) with respect to saline controls. All error bars indicate s.e.m. Olig, oligomycin; Palm, palmitate; Pyr/Mal, pyruvate and malate.

## UCP2 mediates ghrelin-induced food intake

To determine whether the UCP2-dependent cellular responses described above are relevant for behaviour, we analysed the feeding behaviour of wild-type and *Ucp2*<sup>-/-</sup> mice in response to ghrelin administration (intraperitoneal, 10 nmol) at the beginning of the dark phase (Fig. 2e). At 1, 2 and 3 h after injection, ghrelin-induced hyperphagia was severely attenuated in *Ucp2*<sup>-/-</sup> mice compared to wild-type mice (1 h: wild type 4.45 ± 0.24, *Ucp2*<sup>-/-</sup> 2.23 ± 0.19, *P* < 0.01; 2 h: wild type 4.41 ± 0.36, *Ucp2*<sup>-/-</sup> 1.72 ± 0.03,



**Figure 2 | Effects of ghrelin on neuronal firing, synaptic plasticity and feeding are mediated by UCP2.** **a**, UCP2 potentiates NPY c-fos activation in the arcuate nucleus in response to ghrelin (10 nmol). Stereological analysis of total NPY-GFP/c-fos cells in the arcuate nucleus of wild-type (*n* = 6) and *Ucp2*<sup>-/-</sup> mice (*n* = 6) is shown. **b**, Ghrelin increases mean action potential frequency of NPY-GFP neuronal perikarya in whole-cell recordings from wild-type mice (*n* = 7); this effect is diminished in *Ucp2*<sup>-/-</sup> mice (*n* = 7). **c**, Ghrelin increases the frequency of mIPSCs on POMC-GFP perikarya in wild type but not in *Ucp2*<sup>-/-</sup> mice (*n* = 5 for both groups). **d**, Ghrelin increases the number of symmetric, putatively inhibitory perikaryal synapses on POMC neurons in wild-type mice but not in *Ucp2*<sup>-/-</sup> mice. **e**, Peripheral ghrelin administration (intraperitoneal) significantly (*P* < 0.05) enhances food intake relative to saline controls in both wild-type (*n* = 6) and *Ucp2*<sup>-/-</sup> (*n* = 6) mice at 1, 2 and 3 h after injection. However, food intake in *Ucp2*<sup>-/-</sup> mice is significantly attenuated compared to wild-type mice 1, 2 and 3 h after injection. **f**, Mediobasal hypothalamus ghrelin injection (500 pmol μl<sup>-1</sup>) significantly increases food intake of both genotypes at 1, 2 and 3 h after injection. However, *Ucp2*<sup>-/-</sup> mice display severely attenuated food intake at all time points compared to wild-type controls. Asterisk indicates statistically significant differences (*P* < 0.05) with respect to saline controls; dagger indicates statistically significant (*P* < 0.05) reduction with respect to ghrelin-treated wild-type mice. All error bars indicate s.e.m.

*P* < 0.001; Fig. 2e, units are fold intake relative to saline controls). We and others recently demonstrated that ghrelin can act at multiple sites in the brain to induce food intake<sup>3,21</sup>. Thus, we injected ghrelin (500 pmol μl<sup>-1</sup>) locally into the mediobasal hypothalamus or ventral tegmental area (VTA) (Supplementary Fig. 4e, f) immediately before the dark phase. At all time points, intra-mediobasal hypothalamus ghrelin-induced food intake was significantly lower in *Ucp2*<sup>-/-</sup> mice compared to wild type (Fig. 2f). Intra-VTA, ghrelin-induced food intake was also attenuated in *Ucp2*<sup>-/-</sup> mice compared to wild type at 1 h but not at 2 or 3 h after ghrelin injection (Supplementary Fig. 4g). These data suggest that UCP2 regulates ghrelin-induced feeding behaviour in both the mediobasal hypothalamus and VTA. The longer lasting effect of ghrelin in the mediobasal hypothalamus suggests that mediobasal hypothalamus control dominates ghrelin-induced feeding behaviour and is consistent with the finding that mice with conditional knockout of arcuate nucleus NPY/AgRP neurons in adulthood do not respond to ghrelin<sup>24</sup>.

We also tested the effect of leptin-reduced food intake in *Ucp2*<sup>-/-</sup> mice compared to wild-type controls. No significant effects were observed between the two groups (Supplementary Fig. 5a, b).

## Ghrelin induces AMPK and ACC phosphorylation

Ghrelin increases AMPK activity within the hypothalamus<sup>25</sup>, including NPY/AgRP neurons<sup>26</sup>, which promotes feeding<sup>25,27</sup> and enhances mitochondrial biogenesis and function in the periphery<sup>28,29</sup>. We found that ghrelin promotes AMPK phosphorylation (pAMPK/AMPK ratio) in a GHSR-dependent manner (Fig. 3a) in both wild-type and *Ucp2*<sup>-/-</sup> mice (Fig. 3b), indicating that ghrelin's action on AMPK is downstream of GHSR but upstream of UCP2. We also observed that ghrelin elevates phosphorylated acetyl-CoA carboxylase (ACC), consistent with increased AMPK activity (Fig. 3b). Intracerebroventricular injection of a stimulator of AMPK activity, 5-aminoimidazole-4-carboxamide-1-β-D-ribofuranoside (AICAR), like ghrelin, increased both mitochondrial coupled and uncoupled respiration in the hypothalamus of wild-type but not *Ucp2*<sup>-/-</sup> mice (Supplementary Fig. 6a). In addition, AICAR promoted food intake in both wild-type and *Ghsr*<sup>-/-</sup> mice in a similar manner (Fig. 3c), whereas AICAR-induced food intake was significantly diminished in *Ucp2*<sup>-/-</sup> mice compared to wild type (Fig. 3d). Intracerebroventricular injection of compound C, an inhibitor of AMPK, inhibited food intake induced after peripheral ghrelin injections in wild-type mice, but did not further decrease food intake in *Ucp2*<sup>-/-</sup> animals (Fig. 3e), and, it also prevented ghrelin-induced elevations in *Ucp2* mRNA levels in wild-type mice (Fig. 3f). These results confirm that ghrelin activates AMPK signalling through the GHSR, and reveal that AMPK requires UCP2 to exert an effect on feeding.

## Ghrelin-induced food intake is mediated by CPT1

AMPK activation and ACC inhibition lead to a reduction in malonyl-CoA and release of carnitine palmitoyltransferase 1 (CPT1) from allosteric inhibition. Increased CPT1 activity enhances oxidation of long chain fatty acids (LCFA) and increases food intake<sup>30</sup>. Ghrelin increased the expression of hypothalamic *Cpt1* mRNA (Fig. 3g) and this was dependent on AMPK activity, as intracerebroventricularly administered compound C blocked this effect (Fig. 3g). The induction of *Cpt1* mRNA by ghrelin occurred rapidly, after just 1 h (Supplementary Fig. 6b). Ghrelin-triggered elevation in *Ucp2*, *Npy* and *Nrf1* mRNA levels also occurred after 1 h (Supplementary Fig. 6c). The ghrelin-induced increase in *Ucp2* mRNA levels was diminished by concomitant intracerebroventricular administration of either compound C or etomoxir (CPT1 inhibitor) (Fig. 3f), confirming that ghrelin-induced *Ucp2* mRNA expression is mediated by the AMPK-CPT1 pathway. Furthermore, etomoxir and compound C prevented the ghrelin-induced increase in both *Npy* and *AgRP* gene expression (Fig. 3h and Supplementary Fig. 6d), mimicking the impairments found in *Ucp2*<sup>-/-</sup> mice (see Fig. 1e and

Supplementary Fig. 1f). No difference in *Pomc* mRNA was observed in any group (data not shown). In line with the mRNA expression data, etomoxir also suppressed ghrelin-induced food intake (Fig. 3i).

The aforementioned data indicate that CPT1 activation, similar to AMPK activation, is upstream of UCP2. However, we found that ghrelin-induced elevation of hypothalamic *Cpt1* mRNA levels is not observable in *Ucp2*<sup>-/-</sup> animals (Fig. 3j). This apparent paradox is resolved in studies described below regarding hypothalamic free radicals.

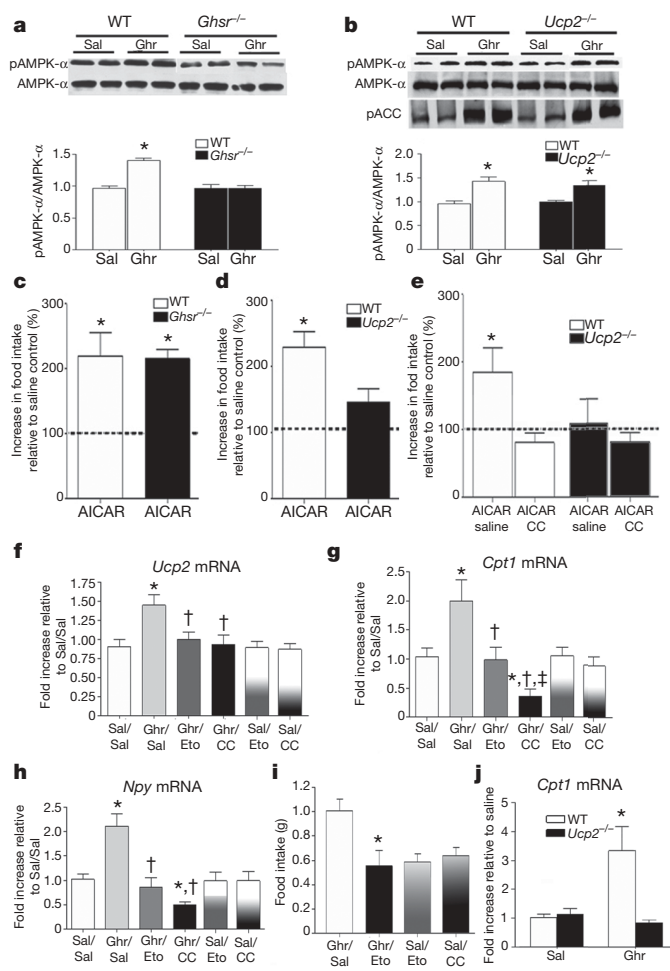
### Ghrelin raises hypothalamic LCFA CoAs and blood NEFAs

To assess whether ghrelin increases the availability of substrates for mitochondrial beta oxidation coinciding with AMPK and CPT1 activation, we analysed hypothalamic levels of LCFA CoAs 3 h after ghrelin administration. Total LCFA CoAs (comprising both 16- and 18-carbon LCFA CoAs) were significantly higher in the hypothalamus of ghrelin- versus saline-treated controls (saline 29.15 ± 1.17, ghrelin 33.72 ± 1.15 nmol metabolite per g tissue; *P* < 0.05; *n* = 4 and 6 respectively; Supplementary Fig. 6e). Next, we analysed LCFA CoAs in the hypothalamus at 0, 30, 60, 120 and 180 min after ghrelin administration. Compared to time 0 (36.18 ± 2.25 nmol metabolite per g tissue), by 30 min (in the absence of food) hypothalamic LCFA CoA levels were significantly elevated (40.75 ± 1.55 nmol metabolite per g tissue, *P* < 0.05) and remained high through 60 (43.08 ± 0.95 nmol metabolite per g tissue), 120 (42.24 ± 0.62 nmol metabolite per g tissue) and 180 min (40.9 ± 0.42 nmol metabolite per g tissue; Supplementary Fig. 6f). Although no statistical differences were detected between these values, the trend suggests that from 60 min onwards, hypothalamic levels of LCFA CoAs may start to descend (Supplementary Fig. 6f), perhaps due to increased use and oxidation of LCFA CoAs. Supporting this

idea is the paradigm in which chronically elevated ghrelin (20 h of fasting) resulted in depletion of hypothalamic levels of LCFA CoAs<sup>31</sup>. We also found that non-esterified fatty acid (NEFA) levels in the circulation (0.23 ± 0.3 mequiv. l<sup>-1</sup> at time 0) start to rise 30 min after ghrelin injection (0.30 ± 0.3 mequiv. l<sup>-1</sup>, *P* > 0.05), and become significantly higher at 60 (0.42 ± 0.3 mequiv. l<sup>-1</sup>, *P* < 0.01), 120 (0.51 ± 0.07 mequiv. l<sup>-1</sup>) and 1,080 min (1.09 ± 0.11 mequiv. l<sup>-1</sup>, *P* < 0.00001) (Supplementary Fig. 6g). The elevated NEFA levels during negative energy balance accompanied by an initial rise but subsequent decline in hypothalamic LCFA CoA levels is consistent with the conclusions of ref. 31 suggesting that the source of hypothalamic LCFA CoAs is the periphery, and also with our observations that hypothalamic LCFA CoAs may undergo beta oxidation.

### Ghrelin reduces ROS in NPY/AgRP neurons via UCP2

A by-product of increased mitochondrial fatty acid beta oxidation is increased ROS production<sup>32,33</sup>. Ghrelin significantly increased synaptosomal ROS in *Ucp2*<sup>-/-</sup> but not in wild-type hypothalami compared to saline controls (Fig. 4a). Next, we injected ghrelin or saline together with dihydroethidium bromide (DHE), a compound that fluoresces as it is oxidized, to analyse *in situ* ROS levels in NPY-GFP and POMC-GFP mice on wild-type and *Ucp2*<sup>-/-</sup> backgrounds. There was no difference in the amount of fluorescent DHE in NPY neurons of wild-type animals treated with either ghrelin or saline (Fig. 4b and Supplementary Fig. 7). However, ghrelin treatment of *Ucp2*<sup>-/-</sup> animals resulted in increased levels of DHE in NPY neurons compared to saline controls (Fig. 4b and Supplementary Fig. 7). The analysis of randomly selected, non-GFP-positive cells showed a shift in DHE levels similar to that seen in NPY neurons (quantified data not shown). In contrast, however, DHE levels in POMC-GFP neurons were significantly higher in wild-type mice treated with



**Figure 3 | Ghrelin activates an intracellular pathway of fatty acid metabolism.** **a**, Ghrelin increases the pAMPK/AMPK ratio in wild-type but not *Ghhr*<sup>-/-</sup> mice (*n* = 5). Ghr, ghrelin; Sal, saline. **b**, Ghrelin increases the pAMPK/AMPK ratio and phosphorylates downstream ACC in both wild-type and *Ucp2*<sup>-/-</sup> mice, indicating that activation of AMPK is upstream to increased UCP2 activity (*n* = 6). **c–e**, AMPK regulation affects food intake. **c**, AICAR significantly increases food intake relative to saline controls in wild-type and *Ghhr*<sup>-/-</sup> mice. **d**, AICAR-induced food intake is significantly suppressed in *Ucp2*<sup>-/-</sup> mice compared to controls, indicating that UCP2 is required to translate AMPK activity into a complete food intake response (*n* = 5). **e**, AICAR significantly increases cumulative 3-h food intake relative to saline controls; however, this effect is completely blocked with a concurrent intracerebroventricular injection of compound C (CC, 6 μg μl<sup>-1</sup>), an inhibitor of AMPK activity. AICAR did not affect the cumulative 3-h food intake in *Ucp2*<sup>-/-</sup> mice and this was not affected by concurrent intracerebroventricular injection of compound C. **f**, Intraperitoneal ghrelin administration increases hypothalamic *Ucp2* mRNA; however, this increase is blocked by concurrent injection of etomoxir or CC, indicating that activation of the AMPK–CPT1 pathway is important in the regulation of *Ucp2* mRNA expression. **g**, Intraperitoneal ghrelin administration increases hypothalamic *Cpt1* mRNA expression; however, the induction of *Cpt1* mRNA with ghrelin (Ghr) is blocked when the CPT1 inhibitor etomoxir (Eto), or compound C (CC), is injected intracerebroventricularly. **h**, Peripheral ghrelin injection increases hypothalamic expression of *Npy* mRNA, which is completely suppressed by intracerebroventricular injection of etomoxir and CC. Note that in **f–h** mRNA expression was not affected by intracerebroventricular injection of etomoxir or compound C with saline injected intraperitoneally. **i**, Inhibition of CPT1 activity with an intracerebroventricular injection of etomoxir suppresses ghrelin-induced (IP 10 nmol) food intake compared to mice receiving intraperitoneal ghrelin and intracerebroventricular saline (Ghr/Sal). Neither intracerebroventricular etomoxir nor compound C affected food intake in mice treated intraperitoneally with saline (*n* = 5–7). **j**, Ghrelin increased *Cpt1* mRNA expression in wild-type but not in *Ucp2*<sup>-/-</sup> mice (*n* = 6, *P* < 0.05). Asterisk, statistically significant (*P* < 0.05) with respect to saline controls; dagger, statistically significant (*P* < 0.05) with respect to Ghr/Sal; double dagger, statistically significant (*P* < 0.05) with respect to Ghr/Eto. All error bars indicate s.e.m.

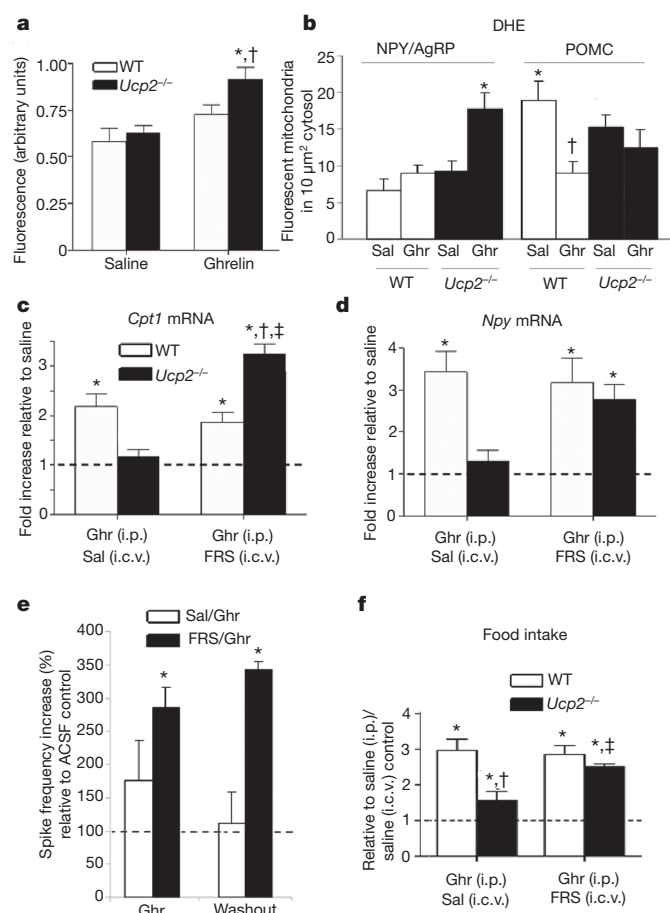
saline compared to those given ghrelin (Fig. 4b and Supplementary Fig. 7). In *Ucp2*<sup>-/-</sup> mice, lower levels of DHE were observed in POMC neurons in both the saline- and ghrelin-treated mice (Fig. 4b and Supplementary Fig. 7). Overall these data are consistent with UCP2's putative ability to buffer ROS levels.

### ROS scavenging reverses responses of *Ucp2*<sup>-/-</sup> mice

In light of the data on ROS, we hypothesized that the paradoxical lack of *Cpt1* mRNA induction by ghrelin in *Ucp2*<sup>-/-</sup> animals (Fig. 3j) may be due to an effect of ROS on CPT1. Thus, we injected a ROS scavenging cocktail containing ascorbic acid (10  $\mu$ M), *N*-acetyl-L-cysteine (10  $\mu$ M) and TEMPO (1  $\mu$ M) intracerebroventricularly into wild-type or *Ucp2*<sup>-/-</sup> mice with a concurrent intraperitoneal injection of ghrelin

(10 nmol). Scavenging of ROS enabled ghrelin to induce elevated *Cpt1* mRNA levels in *Ucp2*<sup>-/-</sup> mice (Fig. 4c).

Our results support the hypothesis that ghrelin-triggered ROS production promotes UCP2 activity and mRNA expression and that the main action of UCP2 in these circuits is to promote ROS scavenging. In line with this, we found that after intracerebroventricular injection of ROS scavenging cocktail, ghrelin-induced elevations in hypothalamic *Ucp2* mRNA levels in wild-type mice were diminished (Supplementary Fig. 8a), which was paralleled by rescue of induction of *Npy* and *AgRP* mRNA levels by ghrelin in *Ucp2*<sup>-/-</sup> mice (Fig. 4d and Supplementary Fig. 8b). In addition, incubating isolated hypothalamic synaptosomes from ghrelin-treated wild-type mice with the ROS scavenging cocktail suppressed ghrelin-induced increases in mitochondrial uncoupled respiration after oligomycin treatment but not in state 3 or FCCP-driven respiration (Supplementary Fig. 8c). When hypothalamic slices from *Ucp2*<sup>-/-</sup> mice were incubated (for 20 min) with the ROS scavenging cocktail, ghrelin-evoked NPY neuronal firing was also restored to wild-type values (Figs 2b and 4e). Moreover, in the presence of the ROS scavenger, washout of ghrelin did not silence NPY neurons (Fig. 4e), indicating that the intracellular events triggered by ghrelin not only potentiated but also maintained neuronal firing subsequent to the initial ghrelin signalling. In association with the effect of ROS scavenger on cellular events in the hypothalamus of *Ucp2*<sup>-/-</sup> mice, ghrelin's ability to promote feeding was restored as well in *Ucp2*<sup>-/-</sup> mice pre-treated with the ROS scavenger (Fig. 4f).



**Figure 4 | ROS is a critical regulator of cellular and behavioural responses to ghrelin.** **a**, Ghrelin increases hypothalamic synaptosomal ROS production in *Ucp2*<sup>-/-</sup> mice but not wild-type mice ( $n = 4-5$ ). **b**, Quantification of ROS production shows that ghrelin markedly increases *in situ* ROS in NPY neurons in *Ucp2*<sup>-/-</sup> but not wild-type mice. **c**, Suppressed *Cpt1* mRNA expression in ghrelin-treated *Ucp2*<sup>-/-</sup> mice after 3 h is completely reversed and elevated above wild-type levels by concurrent intracerebroventricular injection of free radical scavenging (FRS) cocktail ( $n = 5$ ,  $P < 0.05$ ). i.c.v., intracerebroventricular; i.p., intraperitoneal. **d**, FRS cocktail reverses suppressed *Npy* mRNA in ghrelin-treated *Ucp2*<sup>-/-</sup> mice. **e**, Incubating arcuate slices with FRS restores suppressed NPY neuronal action potential firing in *Ucp2*<sup>-/-</sup> mice. Neurons were from NPY-GFP/*Ucp2*<sup>-/-</sup> mice. Asterisk, significant to 100% saline controls. **f**, The FRS cocktail reverses the suppressed ghrelin-induced food intake after 1 h in *Ucp2*<sup>-/-</sup> mice ( $n = 9$ ,  $P < 0.05$ ). Except for **e**: asterisk, significant to saline controls; dagger, significant decrease compared to wild-type mice administered ghrelin intraperitoneally and saline intracerebroventricularly; double dagger, significant with respect to *Ucp2*<sup>-/-</sup> mice administered ghrelin intraperitoneally and saline intracerebroventricularly. All error bars indicate s.e.m.

### Discussion

We have identified intracellular events that contribute to appropriate activation of NPY/AgRP neurons in response to ghrelin (Supplementary Fig. 9). Ghrelin induces a rapid increase of NPY/AgRP neuronal firing by activation of GHSR. Ghrelin, via GHSR, also results in AMPK activation that suppresses ACC activity, eliminating the inhibition of CPT1. CPT1 activation can enhance LCFA entry into mitochondria and increase beta oxidation. Fatty acid beta oxidation promotes generation of ROS<sup>32,33</sup>, which together with fatty acids promote UCP2 transcription and activity<sup>34</sup>. UCP2 activity neutralizes ROS<sup>35</sup>, thereby allowing continuous CPT1-promoted fatty acid beta oxidation that enables continuous support of the bioenergetic needs to maintain firing of NPY/AgRP cells. Sustained firing of NPY/AgRP neurons, the efferents of which are GABAergic<sup>36,37</sup> onto POMC neurons, results in activity-dependent synaptic plasticity promoting an organization of increased inhibitory input onto POMC neurons leading to increased feeding. Our results also indicate that fatty acids, which are elevated in the circulation concurrently with ghrelin during fasting<sup>15</sup>, are acutely increased in the hypothalamus in response to ghrelin administration so that they can serve as a continuous fuel supply for NPY/AgRP neurons during negative energy balance. In light of previous work<sup>34</sup>, we predict that the absence of fatty acid beta oxidation such as during increased carbohydrate load or positive energy balance may affect UCP2 activation and lead to a different cellular outcome. In support of this, it was recently reported that elevated glucose levels in animals on a high fat diet promote UCP2 induction, which was paralleled with impaired ability of POMC neurons to sense glucose<sup>38</sup>. The apparent elevated baseline ROS levels in POMC neurons are consistent with the possibility of increased carbohydrate fuel-based activity in these cells at the time of relative satiety (when UCP2 would not be activated), because carbohydrates, and glucose in particular, are major generators of ROS through increasing electron transport at complex 1. Because NPY neurons show lower ROS production than POMC neurons, it is tempting to speculate that the NPY/AgRP neurons are less susceptible to free-radical-induced damage, leading to resistance of orexigenic hypothalamic mechanisms to cellular stress. Conversely, POMC cells may be more vulnerable to free-radical-induced damage and progressive degeneration, which may have relevance to age-related onset of obesity. Because GHSR and UCP2 expression overlaps in other

brain sites, including the midbrain, we predict that ghrelin modulation of other brain functions will also be regulated by UCP2-modulated fatty acid metabolism.

## METHODS SUMMARY

Detailed methodology can be found in the Supplementary Information. In short, tissue preparation, immunocytochemistry, synaptosomal respiration, Real-time PCR, quantitative and qualitative electron microscopic analyses of mitochondria and synapse number, electrophysiology and feeding behaviour analyses were performed as previously described<sup>1–3,9,23,36,39</sup>.

The mitochondrial membrane potential was measured in isolated synaptosomal mitochondrial fractions from the hypothalamus using  $\Delta\Psi_m$ -sensitive TMRE and fluorescent spectrofluorophotometry. Non-esterified fatty acids were measured in plasma using a colorimetric non-esterified fatty acid (NEFA) kit (Wako) following the manufacturer's instructions. LCFA CoAs were measured by mass spectroscopy.

ROS production was quantified using dichlorodihydrofluorescein diacetate (DCF) in mitochondrial fractions and using DHE in fixed tissue.

Received 2 April; accepted 18 June 2008.

Published online 30 July 2008.

- Cowley, M. A. *et al.* The distribution and mechanism of action of ghrelin in the CNS demonstrates a novel hypothalamic circuit regulating energy homeostasis. *Neuron* **37**, 649–661 (2003).
- Diano, S. *et al.* Ghrelin controls hippocampal spine synapse density and memory performance. *Nature Neurosci.* **9**, 381–388 (2006).
- Abizaid, A. *et al.* Ghrelin modulates the activity and synaptic input organization of midbrain dopamine neurons while promoting appetite. *J. Clin. Invest.* **116**, 3229–3239 (2006).
- Kamegai, J. *et al.* Central effect of ghrelin, an endogenous growth hormone secretagogue, on hypothalamic peptide gene expression. *Endocrinology* **141**, 4797–4800 (2000).
- Diano, S. *et al.* Uncoupling protein 2 prevents neuronal death including that occurring during seizures: a mechanism for preconditioning. *Endocrinology* **144**, 5014–5021 (2003).
- Garcia-Martinez, C. *et al.* Overexpression of UCP3 in cultured human muscle lowers mitochondrial membrane potential, raises ATP/ADP ratio, and favors fatty acid vs. glucose oxidation. *FASEB J.* **15**, 2033–2035 (2001).
- Rossmeisl, M. *et al.* Expression of the uncoupling protein 1 from the ap2 gene promoter stimulates mitochondrial biogenesis in unilocular adipocytes *in vivo*. *Eur. J. Biochem.* **269**, 19–28 (2002).
- Wu, Z. *et al.* Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* **98**, 115–124 (1999).
- Coppola, A. *et al.* A central thermogenic-like mechanism in feeding regulation: an interplay between arcuate nucleus T3 and UCP2. *Cell Metab.* **5**, 21–33 (2007).
- Horvath, T. L. *et al.* Brain uncoupling protein 2: uncoupled neuronal mitochondria predict thermal synapses in homeostatic centers. *J. Neurosci.* **19**, 10417–10427 (1999).
- Willesen, M. G., Kristensen, P. & Romer, J. Co-localization of growth hormone secretagogue receptor and NPY mRNA in the arcuate nucleus of the rat. *Neuroendocrinology* **70**, 306–316 (1999).
- Sun, Y., Asnicar, M., Saha, P. K., Chan, L. & Smith, R. G. Ablation of ghrelin improves the diabetic but not obese phenotype of ob/ob mice. *Cell Metab.* **3**, 379–386 (2006).
- Barazzoni, R. *et al.* Ghrelin regulates mitochondrial-lipid metabolism gene expression and tissue fat distribution in liver and skeletal muscle. *Am. J. Physiol. Endocrinol. Metab.* **288**, E228–E235 (2005).
- Tsubone, T. *et al.* Ghrelin regulates adiposity in white adipose tissue and UCP1 mRNA expression in brown adipose tissue in mice. *Regul. Pept.* **130**, 97–103 (2005).
- Tschop, M., Smiley, D. L. & Heiman, M. L. Ghrelin induces adiposity in rodents. *Nature* **407**, 908–913 (2000).
- Horvath, B. *et al.* Uncoupling protein 2 (UCP2) lowers alcohol sensitivity and pain threshold. *Biochem. Pharmacol.* **64**, 369–374 (2002).
- Scott, I. D. & Nicholls, D. G. Energy transduction in intact synaptosomes. Influence of plasma-membrane depolarization on the respiration and membrane potential of internal mitochondria determined *in situ*. *Biochem. J.* **186**, 21–33 (1980).
- Zhang, C. Y. *et al.* Genipin inhibits UCP2-mediated proton leak and acutely reverses obesity- and high glucose-induced beta cell dysfunction in isolated pancreatic islets. *Cell Metab.* **3**, 417–427 (2006).
- Krauss, S., Zhang, C. Y. & Lowell, B. B. A significant portion of mitochondrial proton leak in intact thymocytes depends on expression of UCP2. *Proc. Natl Acad. Sci. USA* **99**, 118–122 (2002).
- Zhang, C. Y. *et al.* Uncoupling protein-2 negatively regulates insulin secretion and is a major link between obesity, beta cell dysfunction, and type 2 diabetes. *Cell* **105**, 745–755 (2001).
- Abizaid, A., Gao, Q. & Horvath, T. L. Thoughts for food: brain mechanisms and peripheral energy balance. *Neuron* **51**, 691–702 (2006).
- Horvath, T. L. & Gao, X. B. Input organization and plasticity of hypocretin neurons: possible clues to obesity's association with insomnia. *Cell Metab.* **1**, 279–286 (2005).
- Pinto, S. *et al.* Rapid rewiring of arcuate nucleus feeding circuits by leptin. *Science* **304**, 110–115 (2004).
- Luquet, S., Phillips, C. T. & Palmiter, R. D. NPY/AgRP neurons are not essential for feeding responses to glucoprivation. *Peptides* **28**, 214–225 (2007).
- Andersson, U. *et al.* AMP-activated protein kinase plays a role in the control of food intake. *J. Biol. Chem.* **279**, 12005–12008 (2004).
- Kohn, D., Sone, H., Minokoshi, Y. & Yada, T. Ghrelin raises  $[Ca^{2+}]_i$  via AMPK in hypothalamic arcuate nucleus NPY neurons. *Biochem. Biophys. Res. Commun.* **366**, 388–392 (2008).
- Minokoshi, Y. *et al.* AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature* **428**, 569–574 (2004).
- Bergeron, R. *et al.* Chronic activation of AMP kinase results in NRF-1 activation and mitochondrial biogenesis. *Am. J. Physiol. Endocrinol. Metab.* **281**, E1340–E1346 (2001).
- Reznick, R. M. & Shulman, G. I. The role of AMP-activated protein kinase in mitochondrial biogenesis. *J. Physiol.* **574**, 33–39 (2006).
- Lam, T. K., Schwartz, G. J. & Rossetti, L. Hypothalamic sensing of fatty acids. *Nature Neurosci.* **8**, 579–584 (2005).
- Wolfgang, M. J. *et al.* Regulation of hypothalamic malonyl-CoA by central glucose and leptin. *Proc. Natl Acad. Sci. USA* **104**, 19285–19290 (2007).
- Du, X. *et al.* Insulin resistance reduces arterial prostacyclin synthase and eNOS activities by increasing endothelial fatty acid oxidation. *J. Clin. Invest.* **116**, 1071–1080 (2006).
- Yamagishi, S. I. *et al.* Leptin induces mitochondrial superoxide production and monocyte chemoattractant protein-1 expression in aortic endothelial cells by increasing fatty acid oxidation via protein kinase A. *J. Biol. Chem.* **276**, 25096–25100 (2001).
- Echtay, K. S. *et al.* Superoxide activates mitochondrial uncoupling proteins. *Nature* **415**, 96–99 (2002).
- Brand, M. D. *et al.* Mitochondrial superoxide: production, biological effects, and activation of uncoupling proteins. *Free Radic. Biol. Med.* **37**, 755–767 (2004).
- Cowley, M. A. *et al.* Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature* **411**, 480–484 (2001).
- Horvath, T. L., Bechmann, I., Naftolin, F., Kalra, S. P. & Leranth, C. Heterogeneity in the neuropeptide Y-containing neurons of the rat arcuate nucleus: GABAergic and non-GABAergic subpopulations. *Brain Res.* **756**, 283–286 (1997).
- Parton, L. E. *et al.* Glucose sensing by POMC neurons regulates glucose homeostasis and is impaired in obesity. *Nature* **449**, 228–232 (2007).
- Gao, Q. *et al.* Anorectic estrogen mimics leptin's effect on the rewiring of melanocortin cells and Stat3 signaling in obese animals. *Nature Med.* **13**, 89–94 (2007).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by NIH grants to T.L.H., S.D., X.-B.G. and G.I.S., by a New Zealand Foundation for Research Science and Technology (FRST) fellowship to Z.B.A., by grants from the JDRF and American Diabetes Association to S.D., and by a grant from the Michael J. Fox Foundation to T.L.H. We thank B. Lowell for providing breeding pairs of *Ucp2*<sup>−/−</sup> mice and M. Sleeman for providing breeding pairs of *Ghr*<sup>−/−</sup> mice. S.D. thanks A. Lombardi for the discussion on mitochondrial membrane potential measurements. The authors thank V. Pieribone for the use of a spectrofluorophotometer.

**Author Contributions** Z.B.A., S.D. and T.L.H. designed, executed and performed analysis of experiments and wrote the paper. N.W., D.M.E. and A.C. contributed to the execution of the experiments. M.S. and E.B. contributed to the execution of electron microscopy experiments and analysis of the electron microscopic data. Z.-W.L. carried out electrophysiological recordings. X.-B.G. supervised and analysed the electrophysiological experiments. G.C. and G.I.S. designed, performed and analysed the LCFA CoA and NEFA measurements. J.M.F. and M.H.T. provided critical models and reagents for the study and contributed to the data analyses and discussions.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to T.L.H. (tamas.horvath@yale.edu) or S.D. (sabrana.diano@yale.edu).

## ARTICLES

# Structural basis for translation termination on the 70S ribosome

Martin Laurberg<sup>1\*</sup>, Haruichi Asahara<sup>1\*</sup>, Andrei Korostelev<sup>1\*</sup>, Jianyu Zhu<sup>1</sup>, Sergei Trakhanov<sup>1</sup> & Harry F. Noller<sup>1</sup>

**At termination of protein synthesis, type I release factors promote hydrolysis of the peptidyl-transfer RNA linkage in response to recognition of a stop codon. Here we describe the crystal structure of the *Thermus thermophilus* 70S ribosome in complex with the release factor RF1, tRNA and a messenger RNA containing a UAA stop codon, at 3.2 Å resolution. The stop codon is recognized in a pocket formed by conserved elements of RF1, including its PxT recognition motif, and 16S ribosomal RNA. The codon and the 30S subunit A site undergo an induced fit that results in stabilization of a conformation of RF1 that promotes its interaction with the peptidyl transferase centre. Unexpectedly, the main-chain amide group of Gln 230 in the universally conserved GGQ motif of the factor is positioned to contribute directly to peptidyl-tRNA hydrolysis.**

Protein synthesis is carried out by the ribosome, where genetic information encoded in mRNA is translated into a polypeptide chain. Translation ends when a stop codon on mRNA reaches the decoding centre in the A site of the small ribosomal subunit. In contrast to the recognition of sense codons by tRNAs, stop codons are recognized by extra-ribosomal proteins called class I release factors, which also promote hydrolysis of the peptidyl-tRNA linkage in the peptidyl transferase centre of the large ribosomal subunit<sup>1–3</sup>. In bacteria, the UAG and UAA stop codons are recognized by release factor RF1, and the UGA and UAA codons by RF2 (refs 4–6). The low error frequency ( $10^{-3}$ – $10^{-6}$ ) of stop codon recognition is accomplished without a proof-reading mechanism<sup>7</sup>. In principle, stop codons could be recognized directly by specific molecular interactions with the release factors, indirectly by release-factor-promoted interactions with the ribosome, or by some combination of the two. Similarly, hydrolysis of the peptidyl-tRNA linkage could be catalysed by the factors themselves, or by conversion of the ribosomal peptidyl transferase to an esterase by interactions with the factors. Low-resolution cryo-electron microscopy and X-ray studies of termination complexes<sup>8–10</sup> have revealed that the release factors bind to the ribosome in an orientation that roughly coincides with that of a tRNA bound to the A site, placing their extremities in the decoding centre of the small subunit and in the peptidyl transferase centre in the large subunit.

On the basis of mutational studies, a direct recognition model involving a 'tripeptide anticodon' has been proposed for stop-codon recognition<sup>11,12</sup>. According to this model, the conserved motifs PxT (in RF1) or SPF (in RF2) decipher the corresponding stop codons. X-ray structures of the release factors bound to the ribosome at 5.9 Å and 6.7 Å resolution<sup>8</sup> showed that these tripeptide motifs are indeed positioned close to the stop codon, but left the mechanism of recognition unresolved. Although structural changes involving the universally conserved nucleotides guanosine (G)530, adenosine (A)1492 and A1493 of 16S have been shown to have a critical role in discrimination of cognate tRNAs<sup>13–15</sup>, it is unclear whether these nucleotides play an active part in translation termination<sup>16,17</sup>.

Mutational analysis and molecular simulation studies have implicated the universally conserved GGQ motif in hydrolysis of the peptidyl-tRNA ester linkage<sup>18–22</sup>. Although its glutamine side chain has been suggested to be involved directly in catalysis<sup>22,23</sup>, mutation of

this residue has generally led to only small effects<sup>19,21,23,24</sup>, whereas mutation of the glycines confers up to  $10^4$ -fold slower rates of peptide release<sup>18,21</sup>. Proposed roles for the GGQ motif include orienting a water molecule for nucleophilic attack<sup>22,23</sup>, passively opening a path for access of a water molecule to the peptidyl transferase centre (PTC)<sup>25</sup> and excluding other nucleophiles from the esterase reaction<sup>21</sup>. Finally, the mechanism by which peptidyl-tRNA hydrolysis is coupled to stop-codon recognition is not understood. Thus, the molecular bases of all three of the principal mechanisms of translation termination remain unresolved.

Here we present the 3.2 Å crystal structure of a translation termination complex containing the *T. thermophilus* 70S ribosome, a defined mRNA, an initiator tRNA<sup>fMet</sup> bound to an AUG codon in the P site, non-cognate tRNA<sup>fMet</sup> in the E site and release factor RF1 bound in response to a UAA stop codon in the A site. The structure reveals that during termination the proline and threonine side chains of the PxT motif both contact the stop codon bases, accompanied by conformational changes in the decoding centre. However, discrimination of the stop codon also involves a network of interactions between the mRNA, 16S rRNA, and backbone and side-chain atoms of RF1 outside the PxT motif. Although Gln 230 of the conserved GGQ motif interacts directly with the PTC, its main-chain amide group, rather than its side chain, is positioned to play a critical part in catalysis. Our structure suggests that stop-codon recognition is coupled to rearrangement of a loop between domains 3 and 4 of RF1, which results in positioning of the GGQ motif in the PTC.

## Recognition of the stop codon

The overall shape and orientation of the release factor in the 70S ribosome (Fig. 1) is similar to that reported in low-resolution X-ray and cryo-electron microscopy studies<sup>8–10</sup>. Domains 2, 3 and 4 of the factor occupy the ribosomal A site, superimposing approximately on the position of tRNA bound to the A site (Fig. 1a, b). The conformations of the decoding centre of the ribosome and of the stop codon in the termination complex differ markedly from those in sense-codon-recognition complexes<sup>14,15</sup> (Supplementary Fig. 2b). In the sense-codon complex, A1492 and A1493 of 16S rRNA both flip out from their ground-state positions in helix 44 (h44) to form A-minor interactions with the minor groove of the codon-anticodon

<sup>1</sup>Department of Molecular, Cell and Developmental Biology and Center for Molecular Biology of RNA, University of California at Santa Cruz, Santa Cruz, California 95064, USA.

\*These authors contributed equally to this work.

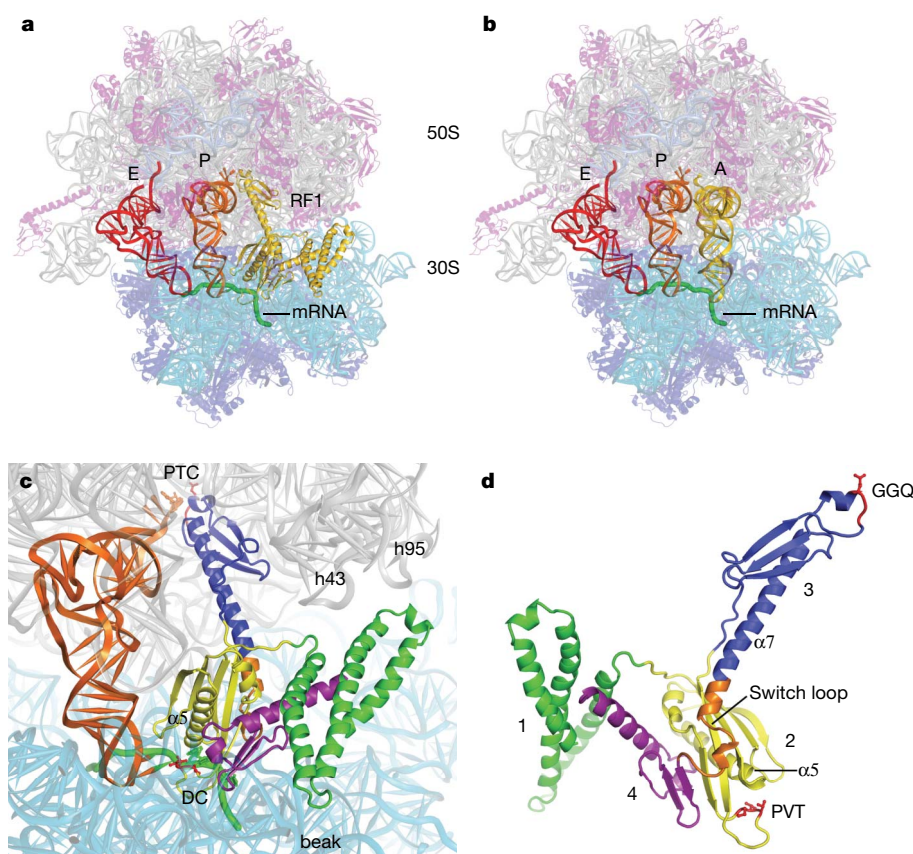
helix (Supplementary Fig. 2b)<sup>14,15</sup>. In the termination complex, only A1492 flips out, leaving A1493 stacked within h44 (Fig. 2a and Supplementary Fig. 2d). A1913, at the tip of the loop of h69 of the 23S rRNA, occupies the space vacated by A1492 in h44 and stacks on A1493 (Fig. 2a and Supplementary Fig. 2d). This conformational change seems to have a critical role in the termination mechanism, as explained later.

The UAA stop codon is bound in a pocket formed by conserved elements of 16S rRNA and domain 2 of RF1, which contains the proposed PxT tripeptide anticodon motif<sup>12</sup> (Fig. 2 and Supplementary Figs 3 and 4). Packing of the tip of helix  $\alpha 5$  of RF1 at Gly 116 against the Watson–Crick edge of uridine (U)1 of the stop codon (Fig. 2b) discriminates against a purine at this position. A hydrogen bond between U1 and the backbone of Gly 116 is possible only with U in this position, as are hydrogen bonds from the Glu 119 backbone amide and the hydroxyl of Thr 186 in the PxT motif to the O4 position of U1. These extensive interactions explain why discrimination is strongest for the first base of the stop codon<sup>7</sup>. In the second position, A2 of the stop codon is stacked between U1 on one side and the imidazole of His 193 on the other. The Watson–Crick edge of A2 packs against the side-chains of Pro 184 and Glu 119; its 6-amino (N6) group hydrogen bonds with the hydroxyl of Thr 186 (Fig. 2b and Supplementary Fig. 4a). Discrimination against guanosine (G) at the second position is probably explained by the inability of its O6 to hydrogen bond with Thr 186 (which donates its only hydrogen to U1). It is less obvious why pyrimidines are excluded here; possible explanations are a weaker propensity for stacking with U1 and the energy penalty caused by a loss of packing against the hydrophobic moieties of Pro 184 and Glu 119. Thus, the PxT motif interacts with the first two bases of the stop codon, rather than with the second and

third bases, as proposed previously<sup>11,12</sup>; however, this motif is by no means the sole determinant of codon recognition.

A3 is unstacked from the first two bases of the stop codon, diverted by the RF1 main chain at residues 192–193 into a position that differs markedly from that of the third base in a sense codon<sup>14,15</sup> (Fig. 2c), stacking instead on G530. Its N6 amino group donates hydrogen bonds to the side chains of Gln 181 and Thr 194, and its N7 accepts a hydrogen bond from the side chain of Thr 194 (Fig. 2c), permitting recognition of both A and G by RF1, while discriminating against pyrimidines. In the backbone of the stop codon, the 2'-hydroxyls of U1 and A2 interact with phosphate 1493 and ribose 1492, respectively, while the 2'-hydroxyl of A3 hydrogen bonds with the main-chain carbonyl of Ile 192 (Fig. 2b, c). Surprisingly, substituting deoxyriboses in the backbone of the stop codon confers only marginal defects in peptide release and RF1 binding<sup>17</sup>, suggesting that these interactions are not critical for stop-codon recognition.

Our structure explains the recently reported observation of competitive binding between RF1 and paromomycin to the ribosome<sup>17</sup>. When paromomycin is bound, A1493 flips out of h44 (ref. 14), sterically preventing the binding of RF1 (Supplementary Fig. 2c); conversely, in the RF1 complex, A1493 occupies the binding site for paromomycin in h44. RF1 function thus involves participation of nucleotides G530, A1492 and A1493, although in an entirely different way from that seen in sense-codon recognition<sup>16,17</sup>. Superposition of the recognition domain of RF2 on that of RF1, on the basis of the similarities in folding of domain 2, and in particular the specificity loops for the structures of the free forms of both type I release factors and of RF1 in the termination complex, allows us to extend our observations to suggest how RF2 may recognize its UAA and UGA codons (see Supplementary Information and Supplementary Fig. 6).



**Figure 1 | Structure of RF1 in the termination complex.** **a**, Positions of RF1 (yellow), P-site tRNA (orange), E-site tRNA (red) and a mRNA (green) in the 70S ribosome. **b**, An elongation complex<sup>43</sup> showing the position of A-site tRNA (yellow) in the context of the current structure. **c**, Orientation of RF1, showing the P-site tRNA (orange), the peptidyl transferase centre (PTC), the

decoding centre (DC) and helices h43 and h95 of 23S rRNA. **d**, The structure of RF1 in its ribosome-bound conformation, rotated  $\sim 180^\circ$  from **a** and **c**. GGQ and PVT motifs are shown in red, and the switch loop (see text) in orange. The domains of RF1 (1–4) are indicated.

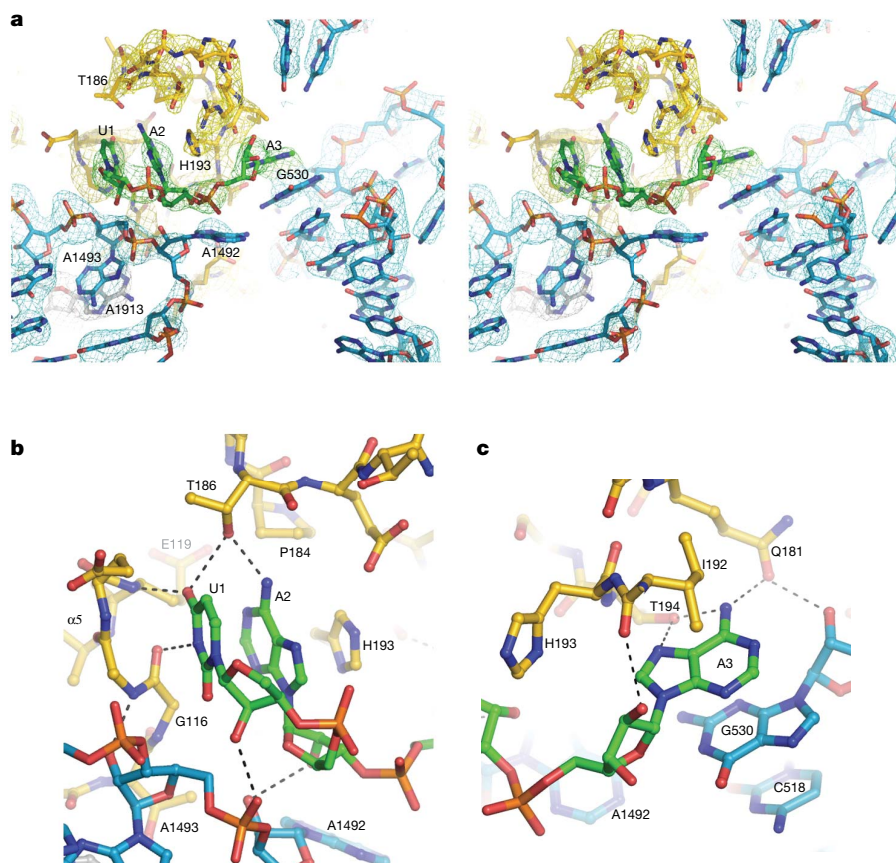
### Peptidyl transferase centre

The GGQ motif at positions 228–230, universally conserved in the type I release factors of all organisms, has been implicated in their ability to promote peptidyl-tRNA hydrolysis<sup>18,20</sup>. Biochemical studies have suggested that the role of the factors in the hydrolysis reaction is to convert the ribosomal peptidyl transferase to an esterase in a reaction using water as an acceptor for the peptidyl moiety<sup>2,3,21</sup>. Our structure shows that the GGQ motif indeed resides in the PTC (Fig. 3a), in agreement with previous low-resolution X-ray findings, although the C $\alpha$ -backbone tracing of this region of RF1 as modelled at 5.9 Å resolution<sup>8</sup> is incompatible with the 3.2 Å electron density map. Residues 228–235 rearrange from their free form<sup>26</sup> to form a short helical element (Fig. 3a, b), which probably contributes to the rigidity of placement of the critical residues. At its end, the GGQ motif interacts with the PTC and with A76 of the P-site tRNA (Fig. 3 and Supplementary Fig. 7). The main-chain amide of Gln 230 is found within hydrogen bonding distance of the 3'OH of A76 of the deacylated tRNA bound in the P site (Fig. 3b), whereas its side-chain amide is positioned away from the location of the scissile ester bond, in a pocket formed by A2451, C2452, U2506 and ribose 76 (Fig. 3b and Supplementary Fig. 7), consistent with its proposed role in discrimination of the attacking nucleophile<sup>23</sup>.

The conformations of nucleotides U2506, G2583, U2584, U2585 and A2602 in the PTC of 23S rRNA are consistent with their sensitivity to the occupancies of the A and P sites<sup>15,25,27</sup> (See Supplementary Information and Supplementary Fig. 8f–h). Mutational analysis has shown A2602 to be critical for release-factor-dependent termination<sup>28–30</sup>. Contrary to proposals concerning its

direct involvement in catalysis of the hydrolysis reaction<sup>22,25,29</sup>, A2602 is buried in a cavity in RF1 and blocked from the catalytic site by the path of the main chain of RF1 at Gly 228 (Fig. 4 and Supplementary Fig. 7c). The observed defects in peptide release by A2602 mutant ribosomes<sup>28–30</sup> must therefore be caused by the contribution of A2602 to the binding and/or positioning of domain 3 of RF1 for catalysis.

Proposed mechanisms for the function of RF1 in peptide release include coordination or activation of the attacking water molecule and induction of conformational changes in the PTC<sup>21–23,30</sup>. In our structure, the main-chain amide of Gln 230 is positioned to form a hydrogen bond with the 3'-hydroxyl group of the terminal nucleoside A76 of the deacylated P-site tRNA. Interpretation at this level of structural detail is facilitated by the location of Gln 230 at the end of a helical element, which, supported by the well-resolved electron density for its side chain, places strong constraints on both the position and the orientation of its backbone amide nitrogen (Supplementary Fig. 7b). Thus, the Gln 230 main-chain amide may help to favour ester hydrolysis by coordinating the leaving group (Fig. 3b, c), analogous to a mechanism that has been proposed for hydrolysis of GTP by Ras<sup>31,32</sup>. Moreover, superposition of the structure of a transition-state analogue bound to the 50S subunit<sup>25,33</sup> with our structure places the oxyanion of the transition-state tetrahedral intermediate at hydrogen-bonding distance from the backbone amide nitrogen of Gln 230 (Fig. 3d), suggesting the possibility of transition-state stabilization, and thus potentially contributing to enhancement of the catalytic rate (Fig. 3d, e). When the structure of an oligonucleotide analogue of deacylated tRNA bound at the A site<sup>34</sup> is superimposed on that of the termination complex, the 3'-hydroxyl of its A76 is also



**Figure 2 | Interactions with the UAA stop codon in the decoding centre of the RF1 termination complex.** **a**, Stereo view of  $\sigma_A$ -weighted  $3F_{\text{obs}} - 2F_{\text{calc}}$  electron density map of the stop codon and surrounding elements of the ribosome and RF1. The density for RF1 was contoured at  $1.0\sigma$ , and for the

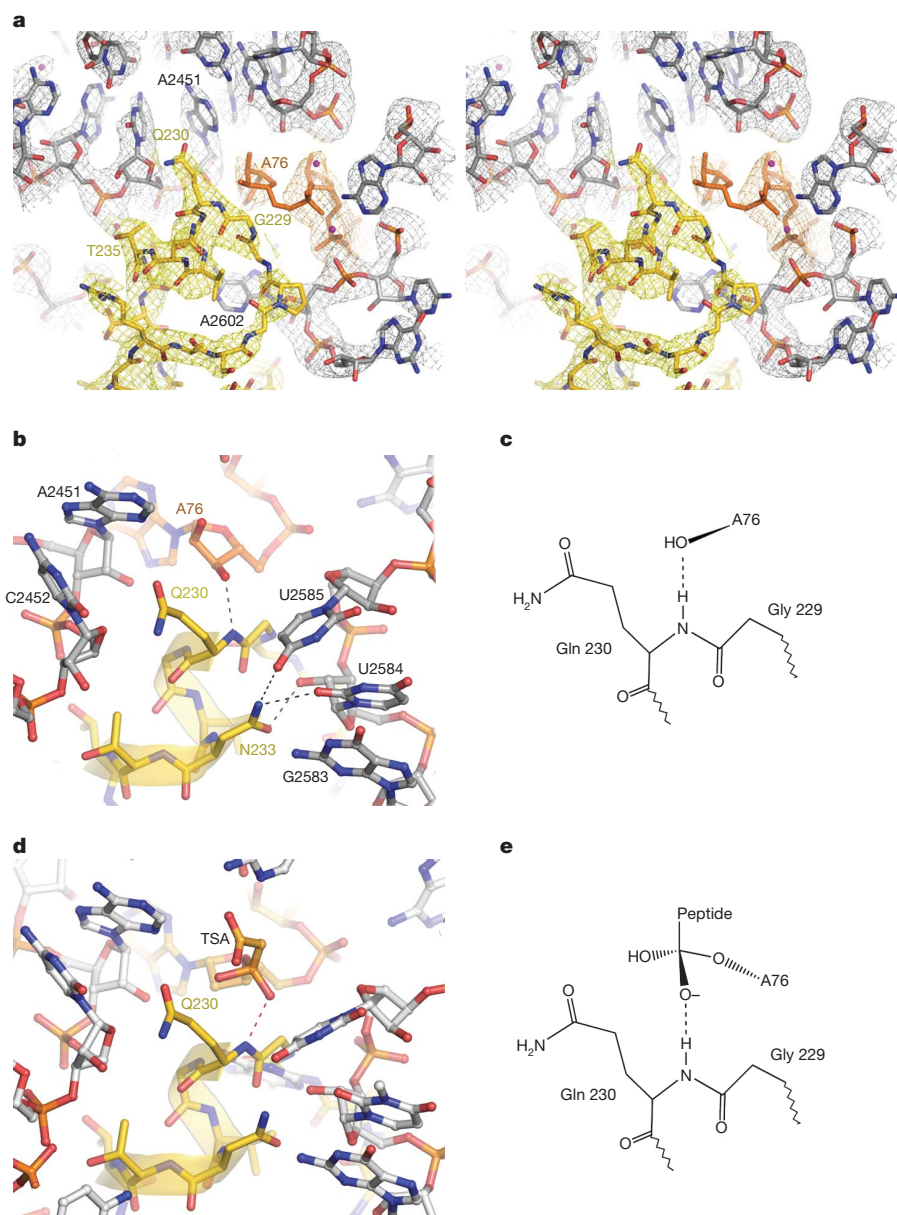
mRNA and rRNA at  $1.5\sigma$ . **b**, Recognition of U1 and A2 of the UAA stop codon. **c**, Discrimination of the third base (A3) by RF1, showing its stacking on G530 of 16S rRNA. 16S rRNA is shown in cyan, RF1 in yellow, mRNA in green and 23S rRNA (A1913) in grey.

found close to the position of the main-chain amide of Gln 230, although it is positioned less optimally for hydrogen bonding with the 3'-hydroxyl of the P-site A76 or with the transition-state analogue. This could explain why binding of a deacylated tRNA to the A site can catalyse peptide release<sup>2</sup>, albeit at lower rates than the class I release factors<sup>35</sup>. Involvement of the main-chain amide group explains why mutation of Gln 230 has only modest effects on catalysis<sup>19,21,24,36</sup>. Conversely, the absence of a side chain in Gly 229 allows the main-chain amide of Gln 230 to approach the 3'-hydroxyl group of A76 of the P-site tRNA and, similarly, the transition-state intermediate, in keeping with the severe defects in peptide release conferred by mutations in Gly 229 of RF1 (refs 18, 21).

### Communication between the decoding site and the PTC

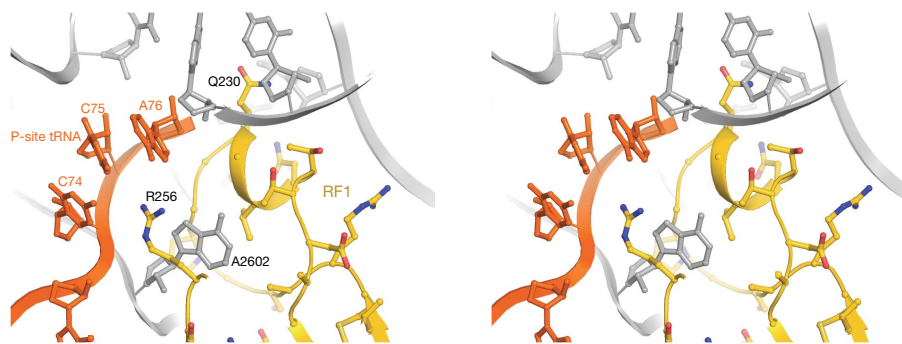
The danger of premature termination is of critical concern in release factor function. Prevention of hydrolysis of the peptidyl-tRNA linkage before recognition of an authentic stop codon requires strict

coordination of these two events. This is probably accomplished by preventing the final docking of RF1 domain 3 in the PTC until codon recognition is established. In the nearly identical crystal structures of isolated RF1 and RF2 (refs 26, 37), the orientation of domains 2 and 3 differs markedly from that observed in their respective termination complexes<sup>8–10</sup>, indicating a propensity for flexibility in their relative positioning. This is supported by direct observation of both the closed<sup>38</sup> and open<sup>39</sup> conformations of free release factors in solution. Comparison of the ribosome-bound RF1 structure with those of free RF1 or RF2 suggests that a key event involves conformational rearrangement of a loop (residues 286–300) connecting domains 3 and 4, which we term the 'switch loop' (Figs 1d and 5). The result is both to relax its tether on domain 3, allowing it to reach the PTC, and to extend  $\alpha 7$  by two helical turns that are formed from loop residues 286–293 (Fig. 5a). On rearrangement, the switch loop is stabilized by interactions within a pocket formed between the ribosome and RF1 (Fig. 5b). The extra turns of helix formed at the end of  $\alpha 7$  (Fig. 5a) are



**Figure 3 | Interactions of the GGQ region of RF1 in the PTC.** **a**, Stereo view of  $\sigma_A$ -weighted  $3F_{\text{obs}} - 2F_{\text{calc}}$  electron density for RF1 (yellow), P-site tRNA (orange) and 23S rRNA (grey) contoured at  $1.7\sigma$ . **b**, Position of Gln 230. **c**, Model for product stabilization by hydrogen bonding between the main-chain amide of Gln 230 and the 3'-OH of A76 of the P-site tRNA.

**d**, Superposition of a peptidyl-transferase transition-state analogue (TSA, orange) complexed with the 50S subunit (grey)<sup>25</sup> on the structure of the termination complex (this work). The main-chain amide of Gln 230 is positioned to hydrogen bond with the oxyanion of the TSA. **e**, Model for transition-state stabilization.



**Figure 4 | Stereo view of the RF1 binding pocket for 23S rRNA nucleotide**

**A2602.** 23S rRNA is shown in grey, P-site tRNA in orange and RF1 in yellow.

held in position by contacts with the  $\beta$ -sheet of domain 2 of RF1 and with residues 1914–1915 in the loop of h69 of 23S rRNA (Fig. 5), directing domain 3 into the PTC.

Rearrangement of the switch loop and positioning of the extended  $\alpha 7$  helix depend on interactions that are made possible by successful stop codon recognition. Flipping out of A1493 or failure of A1913 to stack on A1493 in h44 not only would result in steric clash between these nucleotides and domain 2 of RF1, preventing correct docking of its 'reading head' on the stop codon, but would also affect formation of the binding pocket for the rearranged switch loop and contacts needed to position the extended  $\alpha 7$  helix. Thus, binding of RF1, recognition of the stop codon and the ensuing molecular rearrangements in both the ribosome and the factor leading to docking of the GGQ motif in the PTC are strongly interdependent, cooperative events. The critical interactions involving nucleotides 1913–1915 of 23S rRNA explain the deleterious effects on translation termination caused by deletion of helix 69 of 23S rRNA<sup>40</sup>.

### The mechanism of translation termination

The structure presented here supports the following model for the mechanism of translation termination. The stop codon is bound and recognized in a pocket formed by interactions between the  $\beta$ -sheet of domain 2 of RF1 (including the conserved PxT motif)—the reading head of RF1—and G530, A1492 and A1493—the three critical nucleotides of the 16S rRNA decoding site. A1492 flips out from its position in h44, reminiscent of the events of tRNA selection<sup>14,15</sup>. However, in contrast to the mechanism of sense-codon recognition, A1493 remains stacked within h44, allowing docking of the reading head of RF1 on the stop codon, with which it would otherwise clash. The space vacated by A1492 in h44 is filled by A1913 of 23S rRNA, which stacks on A1493, removing itself and A1493 from a possible clash with RF1. Interactions between RF1 and these rearranged elements of 16S and 23S rRNA induce a conformational change in the switch loop of RF1, which forms an extension of the  $\alpha 7$  helix, creating

a rigid connection between the decoding site and domain 3, docking the latter in the peptidyl transferase centre. Facilitated by product and possibly transition-state stabilization by the main-chain amide group of Gln 230 of RF1, the peptidyl-tRNA linkage is then cleaved. The position of the side chain of the universally conserved Gln 230 is consistent with proposals that it may help to orient the attacking water molecule<sup>22,23</sup> and to discriminate water from other potential nucleophiles<sup>21</sup>, although this would need to be confirmed by the structure of a substrate complex rather than the product complex presented here. An interesting finding is that in the mechanisms of both stop-codon recognition and catalysis, elements of the polypeptide backbone of RF1 seem to have a crucial role. Thus, the basic functions of translation termination could have originated from interactions with simple peptides, possibly as an evolutionary transition from a process that may have initially been carried out by RNA, a possibility supported by the observation that deacylated tRNA bound to the A site is able to catalyse peptide release<sup>2,7,35</sup>.

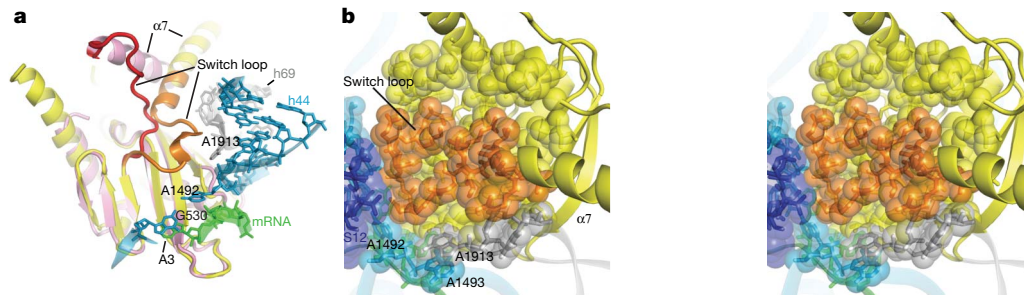
### METHODS SUMMARY

Ribosomes were purified from *T. thermophilus* HB27, and were then dissociated and reassociated to remove endogenous mRNA and tRNAs as described in Supplementary Methods. A 70S termination complex was formed with RF1, a defined mRNA and tRNAs, and was crystallized essentially as described in ref. 15. The crystal structure was solved by molecular replacement and was refined in CNS<sup>41</sup> and PHENIX<sup>42</sup>, as described.

Received 14 April; accepted 23 May 2008.

Published online 2 July 2008.

1. Capecchi, M. R. Polypeptide chain termination *in vitro*: isolation of a release factor. *Proc. Natl Acad. Sci. USA* **58**, 1144–1151 (1967).
2. Caskey, C. T., Beaudet, A. L., Scolnick, E. M. & Rosman, M. Hydrolysis of fMet-tRNA by peptidyl transferase. *Proc. Natl Acad. Sci. USA* **68**, 3163–3167 (1971).
3. Vogel, Z., Zamir, A. & Elson, D. The possible involvement of peptidyl transferase in the termination step of protein biosynthesis. *Biochemistry* **8**, 5161–5168 (1969).
4. Scolnick, E., Tompkins, R., Caskey, T. & Nirenberg, M. Release factors differing in specificity for terminator codons. *Proc. Natl Acad. Sci. USA* **61**, 768–774 (1968).



**Figure 5 | Rearrangements and packing of the switch loop of RF1.**

**a**, Superposition of domain 2 of free RF1 (ref. 37; pink) on that of ribosome-bound RF1 (yellow). Rearrangement of the switch loop (red in free RF1 and orange in bound RF1) results in reorientation of  $\alpha 7$  and extension of its length by two helical turns. Free RF1 is shown because of local disorder in the

structure of the otherwise similar free RF1 (ref. 37). **b**, Stereo view of packing of the switch loop (orange) in a pocket formed by elements of protein S12 (dark blue), the  $\beta$ -sheet of domain 2 of RF1 (yellow), the 530 loop of 16S rRNA (cyan) and nucleotides A1492 of 16S rRNA (cyan) and A1913 of 23S rRNA (grey), which form the floor of the pocket.

5. Capecchi, M. R. & Klein, H. A. Characterization of three proteins involved in polypeptide chain termination. *Cold Spring Harb. Symp. Quant. Biol.* **34**, 469–477 (1969).
6. Scolnick, E. M. & Caskey, C. T. Peptide chain termination. V. The role of release factors in mRNA terminator codon recognition. *Proc. Natl Acad. Sci. USA* **64**, 1235–1241 (1969).
7. Freistroffer, D. V., Kwiatkowski, M., Buckingham, R. H. & Ehrenberg, M. The accuracy of codon recognition by polypeptide release factors. *Proc. Natl Acad. Sci. USA* **97**, 2046–2051 (2000).
8. Petry, S. *et al.* Crystal structures of the ribosome in complex with release factors RF1 and RF2 bound to a cognate stop codon. *Cell* **123**, 1255–1266 (2005).
9. Klaholz, B. P. *et al.* Structure of the *Escherichia coli* ribosomal termination complex with release factor 2. *Nature* **421**, 90–94 (2003).
10. Rawat, U. B. *et al.* A cryo-electron microscopic study of ribosome-bound termination factor RF2. *Nature* **421**, 87–90 (2003).
11. Ito, K., Uno, M. & Nakamura, Y. A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature* **403**, 680–684 (2000).
12. Nakamura, Y. & Ito, K. A tripeptide discriminator for stop codon recognition. *FEBS Lett.* **514**, 30–33 (2002).
13. Ogle, J. M., Carter, A. P. & Ramakrishnan, V. Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem. Sci.* **28**, 259–266 (2003).
14. Ogle, J. M. *et al.* Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897–902 (2001).
15. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
16. Youngman, E. M., Cochella, L., Brunelle, J. L., He, S. & Green, R. Two distinct conformations of the conserved RNA-rich decoding center of the small ribosomal subunit are recognized by tRNAs and release factors. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 545–549 (2006).
17. Youngman, E. M., He, S. L., Nikstad, L. J. & Green, R. Stop codon recognition by release factors induces structural rearrangement of the ribosomal decoding center that is productive for peptide release. *Mol. Cell* **28**, 533–543 (2007).
18. Frolova, L. Y. *et al.* Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* **5**, 1014–1020 (1999).
19. Seit-Nebi, A., Frolova, L., Justesen, J. & Kisselev, L. Class-1 translation termination factors: invariant GGQ minidomain is essential for release activity and ribosome binding but not for stop codon recognition. *Nucleic Acids Res.* **29**, 3982–3987 (2001).
20. Mora, L. *et al.* The essential role of the invariant GGQ motif in the function and stability *in vivo* of bacterial release factors RF1 and RF2. *Mol. Microbiol.* **47**, 267–275 (2003).
21. Shaw, J. J. & Green, R. Two distinct components of release factor function uncovered by nucleophile partitioning analysis. *Mol. Cell* **28**, 458–467 (2007).
22. Trobro, S. & Aqvist, J. A model for how ribosomal release factors induce peptidyl-tRNA cleavage in termination of protein synthesis. *Mol. Cell* **27**, 758–766 (2007).
23. Song, H. *et al.* The crystal structure of human eukaryotic release factor eRF1—mechanism of stop codon recognition and peptidyl-tRNA hydrolysis. *Cell* **100**, 311–321 (2000).
24. Seit Nebi, A., Frolova, L., Ivanova, N., Poltarau, A. & Kisselev, L. Mutation of a glutamine residue in the universal tripeptide GGQ in human eRF1 termination factor does not cause complete loss of its activity. *Mol. Biol. (Mosk.)* **34**, 899–900 (2000).
25. Schmeing, T. M., Huang, K. S., Strobel, S. A. & Steitz, T. A. An induced-fit mechanism to promote peptide bond formation and exclude hydrolysis of peptidyl-tRNA. *Nature* **438**, 520–524 (2005).
26. Vestergaard, B. *et al.* Bacterial polypeptide release factor RF2 is structurally distinct from eukaryotic eRF1. *Mol. Cell* **8**, 1375–1382 (2001).
27. Schuwirth, B. S. *et al.* Structures of the bacterial ribosome at 3.5 Å resolution. *Science* **310**, 827–834 (2005).
28. Amort, M. *et al.* An intact ribose moiety at A2602 of 23S rRNA is key to trigger peptidyl-tRNA hydrolysis during translation termination. *Nucleic Acids Res.* **35**, 5130–5140 (2007).
29. Polacek, N. *et al.* The critical role of the universally conserved A2602 of 23S ribosomal RNA in the release of the nascent peptide during translation termination. *Mol. Cell* **11**, 103–112 (2003).
30. Youngman, E. M., Brunelle, J. L., Kochaniak, A. B. & Green, R. The active site of the ribosome is composed of two layers of conserved nucleotides with distinct roles in peptide bond formation and peptide release. *Cell* **117**, 589–599 (2004).
31. Maegley, K. A., Admiraal, S. J. & Herschlag, D. Ras-catalyzed hydrolysis of GTP: A new perspective from model studies. *Proc. Natl Acad. Sci. USA* **93**, 8160–8166 (1996).
32. Li, G. & Zhang, X. C. GTP hydrolysis mechanism of Ras-like GTPases. *J. Mol. Biol.* **340**, 921–932 (2004).
33. Schmeing, T. M., Huang, K. S., Kitchen, D. E., Strobel, S. A. & Steitz, T. A. Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. *Mol. Cell* **20**, 437–448 (2005).
34. Schmeing, T. M., Moore, P. B. & Steitz, T. A. Structures of deacylated tRNA mimics bound to the E site of the large ribosomal subunit. *RNA* **9**, 1345–1352 (2003).
35. Zavialov, A. V., Mora, L., Buckingham, R. H. & Ehrenberg, M. Release of peptide promoted by the GGQ motif of class 1 release factors regulates the GTPase activity of RF3. *Mol. Cell* **10**, 789–798 (2002).
36. Dincbas-Renqvist, V. *et al.* A post-translational modification in the GGQ motif of RF2 from *Escherichia coli* stimulates termination of translation. *EMBO J.* **19**, 6900–6907 (2000).
37. Shin, D. H. *et al.* Structural analyses of peptide release factor 1 from *Thermotoga maritima* reveal domain flexibility required for its interaction with the ribosome. *J. Mol. Biol.* **341**, 227–239 (2004).
38. Zoldak, G. *et al.* Release factors 2 from *Escherichia coli* and *Thermus thermophilus*: structural, spectroscopic and microcalorimetric studies. *Nucleic Acids Res.* **35**, 1343–1353 (2007).
39. Vestergaard, B. *et al.* The SAXS solution structure of RF1 differs from its crystal structure and is similar to its ribosome bound cryo-EM structure. *Mol. Cell* **20**, 929–938 (2005).
40. Ali, I. K., Lancaster, L., Feinberg, J., Joseph, S. & Noller, H. F. Deletion of a conserved, central ribosomal intersubunit RNA bridge. *Mol. Cell* **23**, 865–874 (2006).
41. Brunger, A. T. *et al.* Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
42. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
43. Yusupov, M. M. *et al.* Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank B. Scott for his input through many stages of this work, and D. Ermolenko, L. Horan, L. Lancaster, B. Scott and D. Staple for comments on the manuscript. We are grateful to the beamline staff at SSRL, ALS and APS for their support during screening and data collection, and to Crystal Chan for her help in using the Berkeley Fermentation Facility. This work was supported by grants from the NIH and NSF (to H.F.N.) and by a fellowship from the Danish Research Council (to M.L.).

**Author Information** Atomic coordinates and structure factors have been deposited with the Protein Data Bank under accession codes 3D5A, 3D5B, 3D5C and 3D5D. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to H.F.N. ([harry@nuvolari.ucsc.edu](mailto:harry@nuvolari.ucsc.edu)).

## LETTERS

# Compositional differences between meteorites and near-Earth asteroids

P. Vernazza<sup>1</sup>, R. P. Binzel<sup>2</sup>, C. A. Thomas<sup>2</sup>, F. E. DeMeo<sup>3</sup>, S. J. Bus<sup>4</sup>, A. S. Rivkin<sup>5</sup> & A. T. Tokunaga<sup>6</sup>

Understanding the nature and origin of the asteroid population in Earth's vicinity (near-Earth asteroids, and its subset of potentially hazardous asteroids) is a matter of both scientific interest and practical importance<sup>1</sup>. It is generally expected that the compositions of the asteroids that are most likely to hit Earth should reflect those of the most common meteorites. Here we report that most near-Earth asteroids (including the potentially hazardous subset) have spectral properties quantitatively similar to the class of meteorites known as LL chondrites. The prominent Flora family in the inner part of the asteroid belt shares the same spectral properties, suggesting that it is a dominant source of near-Earth asteroids. The observed similarity of near-Earth asteroids to LL chondrites is, however, surprising, as this meteorite class is relatively rare (~8 per cent of all meteorite falls). One possible explanation is the role of a size-dependent process, such as the Yarkovsky effect, in transporting material from the main belt.

Spectroscopic observations of more than 400 near-Earth asteroids (NEAs) in visible wavelengths show that 65% of NEAs have S- and Q-type spectral properties<sup>2</sup>. When corrected for discovery biases<sup>3</sup>, the near-Earth population of S- and Q-type asteroids is estimated to be 36% of the total NEA population. Given the dominance of S- and Q-type NEAs and the dominance of ordinary chondrites in meteorite falls (~80%), we focus our analysis on the possible correlations between these two major groups.

To explore the mineralogical composition of potentially hazardous asteroids (PHAs) and the larger population of NEAs, we used the SpeX instrument<sup>4</sup> on the NASA Infrared Telescope Facility (IRTF) on Mauna Kea, Hawaii. This instrument is capable of measuring the flux of reflected sunlight from asteroids over the near-infrared range (0.8–2.5  $\mu\text{m}$ ), which can reveal diagnostic absorption bands due to olivine and pyroxene at 1  $\mu\text{m}$  and pyroxene at 2  $\mu\text{m}$  (Fig. 1). These absorption features are characteristic for both ordinary chondrites and most NEAs.

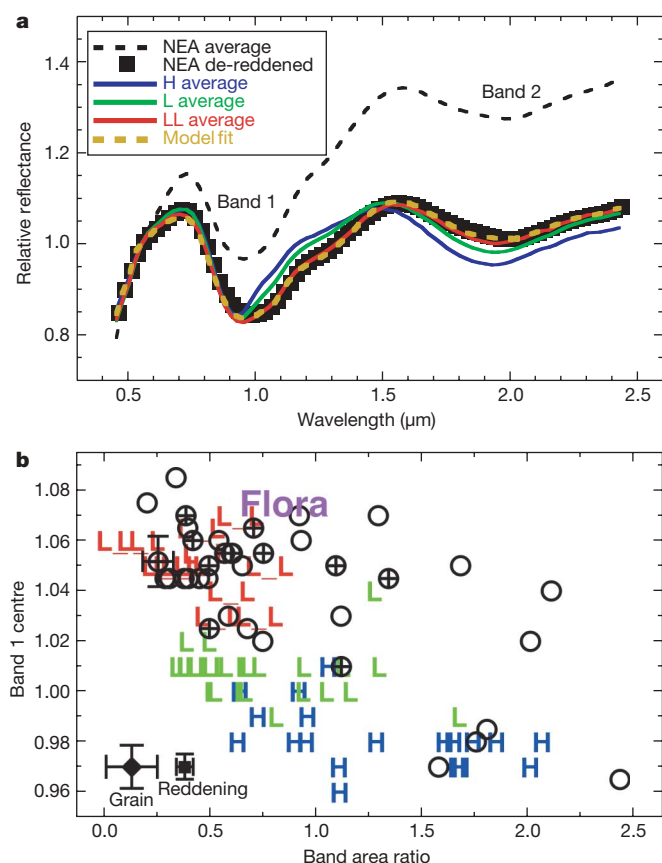
Through a joint observational campaign working cooperatively between MIT, the University of Hawaii and the IRTF (<http://smass.mit.edu/minus.html>), we have collected near-infrared spectra of 38 S- and Q-type NEAs, of which 12 fall into the subset of PHAs, defined as the objects whose trajectories pass within 0.05 AU from the orbit of the Earth. For each of these, the visible wavelength portion of their spectrum was also available from previously published studies<sup>2</sup>. We performed a detailed comparison of these telescopically measured asteroid spectra with analogous wavelength laboratory measurements of ordinary chondrite meteorites catalogued in the RELAB database (<http://www.planetary.brown.edu/rehab/>). Using RELAB measurements for 57 ordinary chondrites, Fig. 1 presents a direct comparison between meteorite and asteroid reflectance spectral properties, including their observational uncertainties.

We applied a radiative transfer model<sup>5</sup> to analyse and compare the mineralogies for both asteroid and meteorite samples. We first applied this model to the ordinary chondrite meteorite spectra to (1) constrain the composition of ordinary chondrites relative to petrographic measurements; and (2) define a robust tracer that would allow us to perform a direct comparison between these meteorites and their possible asteroid parent bodies. We used various end-member minerals found in ordinary chondrites (olivine, orthopyroxene and minor components such as clinopyroxene, chromite and iron). For olivine, orthopyroxene and clinopyroxene, we considered different chemistries, that is, different Mg numbers<sup>6</sup>. In agreement with petrographic measurements, we found that olivine and orthopyroxene are the two main minerals in ordinary chondrites and that clinopyroxene is a minor component (<6%). The inclusion of chromite and iron neither improved our fits nor changed the abundance of olivine, orthopyroxene and clinopyroxene by more than 3%. Thus, we used olivine (ol), orthopyroxene (opx) and clinopyroxene only, and defined as a tracer the ratio of inferred abundances for the two main minerals, ol/(ol+opx). This ratio varies up to ~40% for studied ordinary chondrites. For the three ordinary chondrite subgroups (H, L, LL) we obtain the following average ol/(ol+opx) ratios: H,  $58.8 \pm 4.5\%$ ; L,  $64.2 \pm 6.8\%$ ; and LL,  $75.1 \pm 4.5\%$ . Each of our model-fit average ratios is within 3% of its corresponding value derived from direct laboratory analysis<sup>7</sup>. We find an average ol/(ol+opx) ratio of 63% for all ordinary chondrites in our sample.

We applied the same radiative transfer model<sup>5</sup> to our NEA sample (as well as to members of the Flora family, discussed below). To account for the spectral reddening (if present) due to space weathering processes<sup>8</sup>, we used a space weathering model<sup>9</sup>. We found similar ol/(ol+opx) ratios for both S- and Q-type NEAs, thus demonstrating that space weathering processes do not influence their interpreted mineralogies (see Supplementary Information). We obtain an average ol/(ol+opx) value of 69.6% for all of our S- and Q-type NEA sample. The distribution of inferred ol/(ol+opx) values for studied ordinary chondrites and NEAs are shown in Fig. 2. We weighted each ordinary chondrite subclass (H  $\approx$  43%, L  $\approx$  47% and LL  $\approx$  10%) by their relative fall statistics among all ordinary chondrites.

We find that in both their intrinsic spectral properties (Fig. 1) and in our mineralogical analysis (Fig. 2), our NEA sample does not have the same distribution as the overall population of ordinary chondrite meteorites. On average, NEAs appear more olivine-rich than ordinary chondrites. The best match occurs between the NEAs and the LL ordinary chondrites, where we find that most NEAs (~2/3) give a particularly predominant match to the LL chondrites. We define compatibility between an NEA and an LL chondrite as occurring

<sup>1</sup>Research and Scientific Support Department, European Space Agency, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands. <sup>2</sup>Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. <sup>3</sup>Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique, Observatoire de Paris, 5 Place Jules Janssen, Meudon, F-92195, France. <sup>4</sup>Institute for Astronomy, University of Hawaii, 640 North A'ohoku Place, Hilo, Hawaii 96720, USA. <sup>5</sup>Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA. <sup>6</sup>Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA.

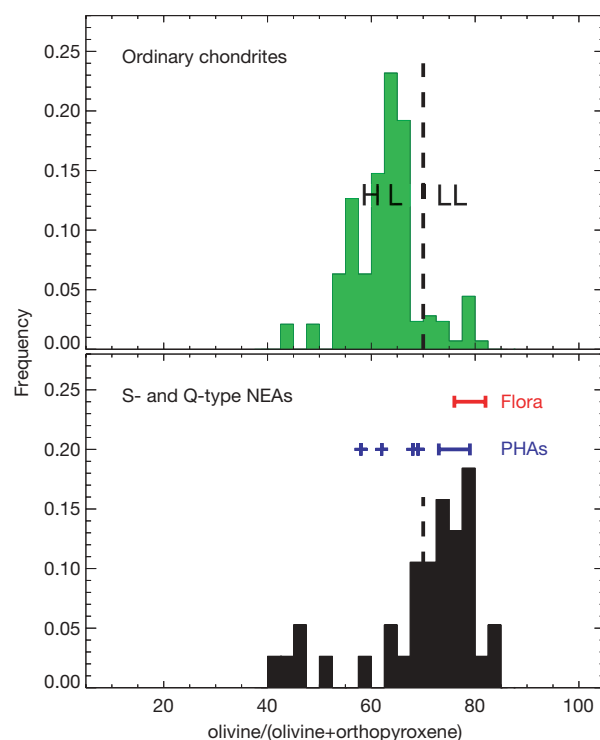


**Figure 1 | Spectral comparisons of asteroids and meteorites.**

**a**, Comparison between the visible to near-infrared spectral signatures of the most common NEAs and ordinary chondrite meteorites. Both populations have similar spectral characteristics, displaying broad absorptions (band 1 and band 2) due to the presence of olivine and pyroxene. For NEAs, these bands are usually superimposed on a spectral continuum whose overall reflectance increases with wavelength. Asteroids showing this reddish continuum are classified as ‘S-types’ while those that match ordinary chondrites (neutral slope) are called ‘Q-types’<sup>19,20</sup>. Space weathering processes similar to those acting on the Moon<sup>8,21</sup> can redden and darken the Q-type spectrum of a fresh asteroid surface, giving it the appearance of an S-type spectrum<sup>2,12,22,23</sup>. The majority (~2/3) of our NEA sample show a band 1 centre at wavelengths >1.02 μm, displayed as an average spectrum before (NEA average) and after (NEA de-reddened) applying a space weathering correction<sup>9</sup>. The average spectra for the H, L and LL ordinary chondrite sub-classes (57 ordinary chondrite spectra in total) and the model fit of the average NEA spectrum ( $ol/(ol+opx) \approx 76\%$ ) are displayed. The majority of NEAs are spectrally similar to, and have modelled mineralogies within the range of, LL chondrite meteorites. **b**, A quantitative comparison and error analysis for the spectral properties for our NEA sample (38 objects; circles) and our meteorite sample (57 objects; H, L, LL symbols). PHAs are inscribed with ‘+’. Band 1 centre is defined as the wavelength that bisects that band into two equal areas. The band area ratio is band 2/band 1, calculated using the modified Gaussian method<sup>24</sup>. (Differences in conventions for fitting the continuum<sup>25</sup> do not allow these ratio values to be directly compared between techniques.) Typical  $1\sigma$  observational error bars (s.d. of the mean) for the asteroids are exemplified by PHA 1862 Apollo (coordinates 0.25, 1.05), based on multiple independent observations. (Meteorite measurements are consistently repeatable in the laboratory and have uncertainties small compared to their plotted symbols.) Spectral band centres and areas are also subject to uncertainties arising from variations in grain size and space weathering; the points at lower left depict modelled uncertainties arising from extreme ranges of these effects. Within the span of all uncertainties, ~2/3 of sampled S- and Q- type NEAs (and equivalently for the PHAs) can be seen to overlap the spectral parameter space of LL chondrite meteorites.

when the  $ol/(ol+opx)$  ratio for the former falls into the LL compositional range (70–85%). This result is surprising, because LL chondrites are the least abundant ordinary chondrites (they represent only 10% of all ordinary chondrites, and 8% of all meteorites). It may suggest that the NEAs we sample telescopically (with radii of 0.3 km to 10 km) are not the immediate parent bodies of smaller objects that fall to Earth as meteorites (that is, pre-atmospheric meteorite parent bodies having radii on the order of metres). Our interpreted mineralogical difference between 0.3–10 km NEAs and ordinary chondrite meteorites suggests that different dynamical mechanisms and/or main-belt source regions may be responsible for supplying these two sample populations (see Supplementary Discussion).

Dynamical simulations predict that ~40% of all roughly kilometre-sized NEAs should be delivered from the asteroid belt to Earth-crossing orbits by the  $v_6$  resonance located on the innermost edge of the main belt<sup>10</sup>. Interestingly, the S-type Flora family, which accounts for 15–20% of the asteroids residing in the inner main-belt region between 2.1 and 2.5 AU, borders the  $v_6$  resonance and is expected to contribute substantially to the population of NEAs<sup>11</sup>. Using our same mineralogical modelling technique to analyse spectra of the asteroid Flora and its fellow family members, we find they have  $ol/(ol+opx)$  values in the 76–82% range. This relatively high olivine abundance of the Flora family agrees with previous measurements<sup>12,13</sup> and is in excellent agreement with the distribution of  $ol/(ol+opx)$  ratios for our NEA sample.



**Figure 2 | Model results for the ratio of olivine to olivine+orthopyroxene for 57 ordinary chondrites and 38 S- and Q-type NEAs.** **a**, Ordinary chondrites; **b**, S- and Q-type NEAs. We also depict the inferred compositional ranges for the Flora family and PHAs (8 of 12 observed PHAs have  $ol/(ol+opx)$  ratios in the range 73–79%; the 4 others are displayed with ‘+’ symbols; error bars are  $1\sigma$ ). About 95% ( $\pm 4\%$ ) of H and L ordinary chondrites are on the left of the dashed line at  $ol/(ol+opx) = 70$  and ~95% ( $\pm 4\%$ ) of LL ordinary chondrites are on the right. The x position of the H, L and LL labels indicates their average  $ol/(ol+opx)$  value. We define compatibility between an NEA and an LL chondrite when the  $ol/(ol+opx)$  ratio for an NEA falls into the LL compositional range (70–85%). It appears that most ordinary chondrites and large S- and Q-type NEAs do not have compatible mineralogy. Most NEAs (~63%) are compatible with an LL mineralogy ( $ol/(ol+opx)$  in the 70–85% range), and account for 10% of the ordinary chondrite falls.

Our results have two direct implications. First, our observed compositional link of LL chondrites with the Flora family appears inconsistent with Flora as the source for shocked L chondrites<sup>14</sup> (~25% of all falls), assuming Flora is a relatively homogeneous parent body. Second, the particular predominance of PHAs (as a representative sample of all NEAs) falling into the LL chondrite category would suggest that a majority of terrestrial craters may have been produced by LL chondrite impactors. At present there are only sparse data showing a mix of L and LL terrestrial impactors<sup>15,16</sup>, but this type of analysis could serve as a geological test of our results.

The present observational results also raise, but do not resolve, two questions. First, why do approximately 2/3 of all NEAs appear to have compositions most similar to a class of meteorites (LL chondrites) that comprise only 8% of all meteorite falls? And second, why does the Flora family (which we suggest dominates the kilometre-sized bodies) not also dominate the flux at the smaller (metre-scale) sizes that produce the greatest number of meteorites? Size range may be the key difference between observed NEAs and the bodies that frequently deliver meteorites (see Supplementary Notes).

Our PHA and NEA sample comprises approximately kilometre-sized bodies, whereas meteorite precursor bodies are most frequently a few metres in size. We note that size is a key factor in the efficiency of the Yarkovsky effect<sup>10,17,18</sup>, which causes bodies (of diameter  $D \leq 20$  km) to drift in orbital semi-major axis and encounter major resonances that speed their delivery to the inner Solar System. At the largest sizes (kilometres), the Yarkovsky drift is expected to be limited and relatively slow (0.01 AU over ~0.5 Gyr), which implies that the source regions of observable PHAs and NEAs most probably border on resonances. Thus the Flora asteroid family, located at the belt's inner edge, adjacent to the highly effective  $v_6$  resonance, is particularly suited as a predominant source for delivering kilometre-sized objects to the vicinity of Earth.

If kilometre-sized objects are indeed predominantly delivered from the inner asteroid belt (and perhaps the Flora family), then Yarkovsky drift and the same  $v_6$  resonance will be even more effective for delivering an inbound flux of potential meteorite bodies in the metre-size range. At such sizes, Yarkovsky drift is predicted to be much more substantial and rapid (0.1 AU over ~50 Myr) throughout the entire inner asteroid belt. Thus metre-sized bodies throughout a much greater breadth of the asteroid belt may be drifted into a number of possible resonances for transport to the inner Solar System. If this is the case, the compositional distribution of meteorite falls is likely to be very different from that of a sample drawn only from the inner edge. Thus a heliocentric-distance-dependent and size-dependent process such as the Yarkovsky effect may account for the apparently different compositional distributions we find between the kilometre-sized and metre-sized populations of bodies intersecting the orbit of Earth.

Received 3 March; accepted 27 May 2008.

1. Gehrels, T., Matthews, M. S. & Schumann, A. M. *Hazards Due to Comets and Asteroids* (Univ. Arizona Press, 1994).
2. Binzel, R. P. *et al.* Observed spectral properties of near-Earth objects: Results for population distribution, source regions, and space weathering processes. *Icarus* **170**, 259–294 (2004).
3. Stuart, J. S. & Binzel, R. P. Bias-corrected population, size distribution, and impact hazard for the near-Earth objects. *Icarus* **170**, 295–311 (2004).
4. Rayner, J. T. *et al.* A medium-resolution 0.8–5.5 micron spectrograph and imager for the NASA Infrared Telescope Facility. *Publ. Astron. Soc. Pacif.* **115**, 362–382 (2003).
5. Shkuratov, Y., Starukhina, L., Hoffmann, H. & Arnold, G. A model of spectral albedo of particulate surfaces: Implications for optical properties of the Moon. *Icarus* **137**, 235–246 (1999).

6. Lucey, P. G. Model near-infrared optical constants of olivine and pyroxene as a function of iron content. *J. Geophys. Res.* **103**, 1703–1714 (1998).
7. Hutchison, R. *Meteorites: A Petrologic, Chemical and Isotopic Synthesis* (Cambridge Univ. Press, 2004).
8. Clark, B. E., Hapke, B., Pieters, C. & Britt, D. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 585–599 (Univ. Arizona Press, 2002).
9. Brunetto, R. *et al.* Modeling asteroid surfaces from observations and irradiation experiments: The case of 832 Karin. *Icarus* **184**, 327–337 (2006).
10. Bottke, W. F., Vokrouhlický, D., Rubincam, D. P. & Brož, M. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 395–408 (Univ. Arizona Press, 2002).
11. Nesvorný, D., Morbidelli, A., Vokrouhlický, D., Bottke, W. F. & Brož, M. The Flora family: A case of the dynamically dispersed collisional swarm? *Icarus* **157**, 155–172 (2002).
12. Chapman, C. R. S-type asteroids, ordinary chondrites, and space weathering: The evidence from Galileo's fly-bys of Gaspra and Ida. *Meteorit. Planet. Sci.* **31**, 699–725 (1996).
13. Gaffey, M. J. Rotational spectral variations of asteroid 8 Flora: Implications for the nature of S-type asteroids and for the parent bodies of ordinary chondrite meteorites. *Icarus* **60**, 83–114 (1984).
14. Nesvorný, D., Vokrouhlický, D., Bottke, W. F., Gladman, B. & Häggström, T. Express delivery of fossil meteorites from the inner asteroid belt to Sweden. *Icarus* **188**, 400–413 (2007).
15. Maier, W. D. *et al.* Discovery of a 25-cm asteroid clast in the giant Morokweng impact crater, South Africa. *Nature* **441**, 203–206 (2006).
16. Tagle, R. & Hecht, L. Geochemical identification of projectiles in impact rocks. *Meteorit. Planet. Sci.* **41**, 1721–1735 (2006).
17. Rubincam, D. P. Asteroid orbit evolution due to thermal drag. *J. Geophys. Res.* **100**, 1585–1594 (1995).
18. Bottke, W. F., Rubincam, D. P. & Burns, J. A. Dynamical evolution of main belt meteoroids: Numerical simulations incorporating planetary perturbations and Yarkovsky thermal forces. *Science* **145**, 301–331 (2000).
19. McFadden, L. A., Gaffey, M. J. & McCord, T. B. Near-earth asteroids — Possible sources from reflectance spectroscopy. *Science* **229**, 160–163 (1985).
20. Bus, S. J. *Compositional Structure in the Asteroid Belt: Results of a Spectroscopic Survey*. Thesis, Massachusetts Inst. Technol. (1999).
21. Pieters, C. M. *et al.* Space weathering on airless bodies: Resolving a mystery with lunar samples. *Meteorit. Planet. Sci.* **35**, 1101–1107 (2000).
22. Chapman, C. R. Space weathering of asteroid surfaces. *Annu. Rev. Earth Planet. Sci.* **32**, 539–567 (2004).
23. Binzel, R. P., Bus, S. J., Burbine, T. H. & Sunshine, J. M. Spectral properties of near-Earth asteroids: Evidence for sources of ordinary chondrite meteorites. *Science* **273**, 946–948 (1996).
24. Sunshine, J. M., Pieters, C. M. & Pratt, S. F. Deconvolution of mineral absorption bands — An improved approach. *J. Geophys. Res.* **95**, 6955–6966 (1990).
25. Gaffey, M. J. *et al.* Mineralogical variations within the S-type asteroid class. *Icarus* **106**, 573–602 (1993).

Supplementary Information is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge S. M. Slivan (Wellesley College) for data processing software development. Most data were acquired by the authors operating as Visiting Astronomers at the IRTF, which is operated by the University of Hawaii under cooperative agreement NNX08AE38A with NASA's Science Mission Directorate, Planetary Astronomy Program. This Letter is based on work supported by the NSF (grant 0506716) and NASA (grant NAG5-12355). Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the NSF or NASA.

**Author Contributions** P.V. performed the quantitative mineral analysis that solidified the results of this Letter, and led the formulation of possible explanations. C.A.T. provided the quantitative analysis of spectral properties. S.J.B. provided the comparison to the Flora region. R.P.B. and A.T.T. served as principal investigators for a joint observing programme to acquire the near-infrared data. Most data were acquired by R.P.B., S.J.B. and C.A.T., while F.E.D. performed most of the processing. Processing routines were developed by S.J.B., A.S.R. and R.P.B. P.V. and R.P.B. worked jointly to write the Letter. All authors discussed the results and commented on the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to P.V. ([pierre.vernazza@esa.int](mailto:pierre.vernazza@esa.int)).

# Testing the speed of 'spooky action at a distance'

Daniel Salart<sup>1</sup>, Augustin Baas<sup>1</sup>, Cyril Branciard<sup>1</sup>, Nicolas Gisin<sup>1</sup> & Hugo Zbinden<sup>1</sup>

Correlations are generally described by one of two mechanisms: either a first event influences a second one by sending information encoded in bosons or other physical carriers, or the correlated events have some common causes in their shared history. Quantum physics predicts an entirely different kind of cause for some correlations, named entanglement. This reveals itself in correlations that violate Bell inequalities (implying that they cannot be described by common causes) between space-like separated events (implying that they cannot be described by classical communication). Many Bell tests have been performed<sup>1</sup>, and loopholes related to locality<sup>2–4</sup> and detection<sup>5,6</sup> have been closed in several independent experiments. It is still possible that a first event could influence a second, but the speed of this hypothetical influence (Einstein's 'spooky action at a distance') would need to be defined in some universal privileged reference frame and be greater than the speed of light. Here we put stringent experimental bounds on the speed of all such hypothetical influences. We performed a Bell test over more than 24 hours between two villages separated by 18 km and approximately east–west oriented, with the source located precisely in the middle. We continuously observed two-photon interferences well above the Bell inequality threshold. Taking advantage of the Earth's rotation, the configuration of our experiment allowed us to determine, for any hypothetically privileged frame, a lower bound for the speed of the influence. For example, if such a privileged reference frame exists and is such that the Earth's speed in this frame is less than  $10^{-3}$  times that of the speed of light, then the speed of the influence would have to exceed that of light by at least four orders of magnitude.

According to quantum theory, quantum correlations violating Bell inequalities simply happen, somehow from outside space-time, in the sense that there is no space-time explanation for their occurrence: there is no event here that somehow influences another distant event there. Yet such a description of correlations, which is radically different from all those found in any other part of science, should be thoroughly tested.

In 1989, Eberhard<sup>7</sup> realized that the existence of a hypothetically privileged reference frame can be experimentally tested for. The idea is that the speed of this influence, although greater than the speed of light, is finite. Hence, if the events are simultaneous in the hypothetically privileged frame, then the signal does not arrive on time and no violation of Bell inequalities should be observed. Note that if the events are simultaneous in some reference frame, then they are also simultaneous with respect to any reference frame moving in a direction perpendicular to the line joining the two events. Accordingly, Eberhard proposed (Ph. H. Eberhard, personal communication) to perform a Bell test over a long distance oriented east–west, over a period of 12 hours (ref. 8). If the events were simultaneous in the Earth's reference frame, then they would also be simultaneous with respect to all frames moving in the plane perpendicular to the east–west axis, and in 12 hours all possible hypothetically privileged frames would be scanned. An intriguing alternative, that we will

not follow here, is to speculate that there is a tachyonic field<sup>9</sup> that couples specifically to entangled particles.

Bohm's<sup>10</sup> pilot-wave model of quantum mechanics is an example of a theory containing an explicit spooky action at a distance, which requires the assumption that there is a universally privileged frame<sup>11</sup>. Also, if the spooky action at a distance propagates at finite speed, then an experiment like the one presented below could possibly falsify the pilot-wave model<sup>12</sup>. Reference 12 stresses that the existence of a universally privileged frame would not contradict relativity.

By 2000, a Bell experiment along the lines presented above had already been analysed<sup>8,13,14</sup>. However, the analysis concerned only two hypothetically privileged reference frames. The first frame was defined by the  $\sim 2.7$  K cosmic microwave background radiation. The second frame analysed was the 'Swiss Alps reference frame', that is, not a universal frame but merely the frame defined by the massive environment of the experiment. The assumption that the privileged frame depends on the experiment's environment leads naturally to the issue of situations in which the massive environments on both sides of the experiment differ, and this was the main subject of the experiment in 2000 (refs 13, 14). In both of these analyses, the hypothetical superluminal influence was termed the speed of quantum information to stress that it is not classical signalling. We shall use this terminology, but we emphasize that this is only the speed of a hypothetical influence and that our result casts very serious doubts on its existence. For views on the speed of quantum information, see ref. 15.

Before presenting our experiment and results, let us clarify the principle of our measurements and how to obtain bounds on this speed of quantum information in any reference frame.

In an inertial reference frame centred on the Earth, two events A and B (in our experiment two single-photon detections) respectively occur at positions  $\mathbf{r}_A$  and  $\mathbf{r}_B$  at times  $t_A$  and  $t_B$ . Let us consider another inertial reference frame F, the hypothetically privileged frame, relative to which the Earth frame moves with a velocity  $\mathbf{v}$  (Fig. 1). When correlations violating a Bell inequality are observed, the speed of quantum information  $V_{\text{QI}}$  in frame F that could cause the correlation is bounded from below as follows:

$$V_{\text{QI}} \geq \frac{|\mathbf{r}'_B - \mathbf{r}'_A|}{|t'_B - t'_A|}$$

where  $(\mathbf{r}'_A, t'_A)$  and  $(\mathbf{r}'_B, t'_B)$  are respectively the coordinates of events A and B in frame F, obtained from  $(\mathbf{r}_A, t_A)$  and  $(\mathbf{r}_B, t_B)$  by Lorentz transformation. After simplification, we obtain

$$\left(\frac{V_{\text{QI}}}{c}\right)^2 \geq 1 + \frac{(1 - \beta^2)(1 - \rho^2)}{(\rho + \beta_{\parallel})^2} \quad (1)$$

where  $\beta = v/c$  is the speed ( $v = |\mathbf{v}|$ ) of the Earth frame in frame F relative to the speed of light ( $c$ );  $\beta_{\parallel} = v_{\parallel}/c$ , with  $v_{\parallel}$  the component of  $\mathbf{v}$  parallel to the A–B axis; and  $\rho = ct_{\text{AB}}/r_{\text{AB}}$  quantifies the alignment of the two events in the Earth frame, with  $t_{\text{AB}} = t_B - t_A$  and  $r_{\text{AB}} = |\mathbf{r}_B - \mathbf{r}_A|$ . In the following, we will consider space-like separated

<sup>1</sup>Group of Applied Physics, University of Geneva, 20 Rue de l'Ecole de Médecine, CH-1211 Geneva 4, Switzerland.

events, for which  $|\rho| < 1$ : the bound on  $V_{\text{QI}}$  in equation (1) will then be larger than  $c$ . For a given privileged frame  $F$ , this bound depends on the orientation of the A–B axis through  $\beta_{\parallel}$  and on the alignment  $\rho$ . To obtain a good lower bound for  $V_{\text{QI}}$ , we should bound the term  $(\rho + \beta_{\parallel})^2$  from above by the smallest possible value, during a period of time  $T$  needed to observe a Bell violation (which, in our experiment, will be the integration time of a two-photon interference fringe).

To gain some intuition, we first consider the simple case in which  $\rho = 0$  (the two events are perfectly simultaneous in the Earth frame) and the A–B axis is perfectly aligned in the east–west direction. Then, when the Earth rotates, there will be a moment  $t_0$  when the east–west direction is perpendicular to  $\mathbf{v}$ , that is,  $\beta_{\parallel}(t_0) = 0$ . During a small time interval around  $t_0$ , we can bound  $|\beta_{\parallel}(t)|$  by a small value, and thus obtain a high lower bound for  $V_{\text{QI}}$ .

In principle, the alignment  $\rho$  could be optimized for each privileged frame that we wish to test, so as to decrease the bound that can be put on  $(\rho + \beta_{\parallel})^2$  during the time interval  $T$  (and increase the term  $(1 - \rho^2)$  at the same time). In our experiment, however, because we want to scan all possible frames, we do not optimize  $\rho$  for each frame; instead, we align the detection events such that  $|\rho| \leq \bar{\rho} \ll 1$ , where  $\bar{\rho}$  is our experimental precision on the alignment  $\rho$ . We then use the fact that  $(1 - \rho^2)/(\rho + \beta_{\parallel})^2 \geq (1 - \bar{\rho}^2)/(\bar{\rho} + |\beta_{\parallel}|)^2$  to get the following bound:

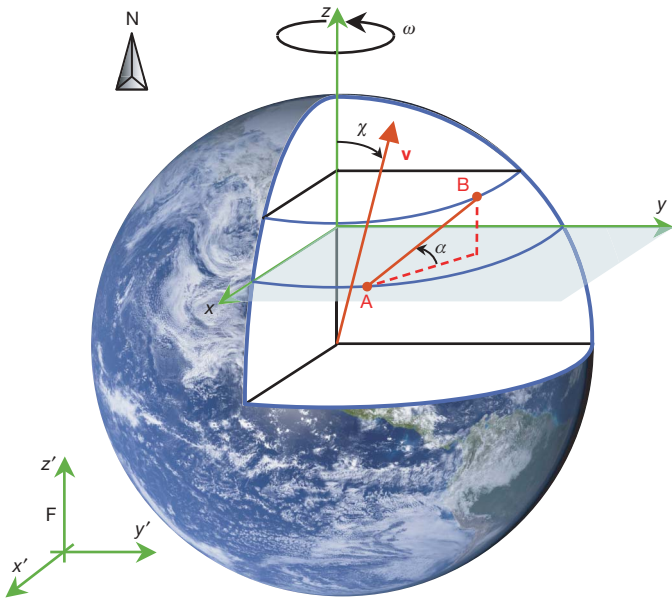
$$\left(\frac{V_{\text{QI}}}{c}\right)^2 \geq 1 + \frac{(1 - \beta^2)(1 - \bar{\rho}^2)}{(\bar{\rho} + |\beta_{\parallel}|)^2} \quad (2)$$

The problem reduces to bounding  $|\beta_{\parallel}|$  directly.

In the configuration of our experiment, the A–B axis is almost, but not perfectly, oriented along the east–west direction. Consequently, the component  $\beta_{\parallel}(t)$  has a 24-hour period, and geometric considerations show that it can be written as (see the Supplementary Information)

$$\beta_{\parallel}(t) = \beta \cos \chi \sin \alpha + \beta \sin \chi \cos \alpha \cos \omega t$$

where  $\chi$  is the zenith angle of  $\mathbf{v}$ ,  $\alpha$  is the angle between the A–B axis and the equatorial  $x$ – $y$  plane (Fig. 1), and  $\omega$  is the angular velocity of the Earth.



**Figure 1 | Reference frames.** The Earth frame moves with respect to a hypothetically privileged reference frame  $F$  at a speed  $\mathbf{v}$ . The zenith angle  $\chi$  between  $\mathbf{v}$  and the  $z$  axis can have values between  $0^\circ$  and  $180^\circ$ . The A–B axis forms an angle  $\alpha$  with the equatorial ( $x$ – $y$ ) plane.  $\omega$ , angular velocity of the Earth.

As we show in the Supplementary Information, in order to bound  $|\beta_{\parallel}|$  from above during a period of time  $T$ , we can consider two cases: one in which  $\mathbf{v}$  does not point close to a pole ( $C_T |\tan \chi| > |\tan \alpha|$ ), and one in which it does ( $C_T |\tan \chi| \leq |\tan \alpha|$ ). Here  $C_T = \cos^2(\omega T/4)$ , which  $\sim 1$  when  $\omega T$  is small. There exists a time interval of length  $T$  during which  $|\beta_{\parallel}(t)|$  is bounded from above by

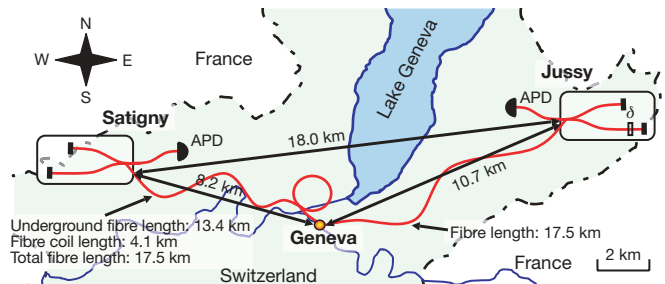
$$|\beta| \sqrt{\sin^2 \chi \cos^2 \alpha - \cos^2 \chi \sin^2 \alpha} \frac{\omega T}{2}$$

in the first case and by

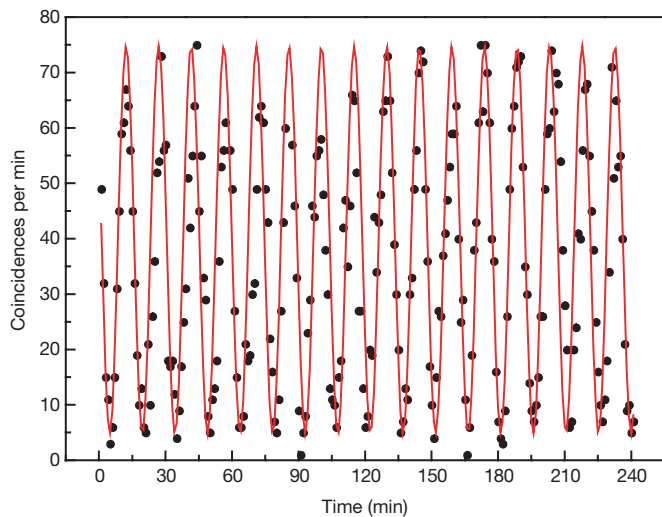
$$|\beta| \left( |\cos \chi \sin \alpha| - |\sin \chi \cos \alpha| \cos \frac{\omega T}{2} \right)$$

in the second case. These bounds, together with equation (2), provide the desired lower bound for  $V_{\text{QI}}$ .

We now describe our experiment. In essence it is a large Franson interferometer<sup>16</sup>. A source situated in our laboratory in Geneva emits entangled photon pairs using the standard parametric down-conversion process in a nonlinear crystal (here a continuous-wave laser pumps a waveguide in a periodically poled lithium niobate crystal)<sup>17</sup>. Using fibre Bragg gratings and optical circulators, each pair is deterministically split and one photon is sent through the Swisscom fibre optic network to Satigny, a village west of Geneva, and the other photon is sent to Jussy, a village east of Geneva. The two receiving stations, located in those two villages, are separated by a direct distance of 18.0 km (Fig. 2). We use energy–time entanglement, a form of entanglement well suited to quantum communication in standard telecommunications fibres<sup>18</sup>. At each receiving station, the photons pass through identically unbalanced fibre-optic Michelson interferometers. The imbalance ( $\sim 25$  cm) is larger than the single-photon coherence length ( $\sim 2.5$  mm), meaning that any single-photon interference is avoided, but is much smaller than the pump laser coherence length ( $> 20$  m). Accordingly, when a photon pair is detected simultaneously in Satigny and Jussy, there is no information about which path—the long arm or the short arm—the photons took in their interferometers. But because the photons were also emitted simultaneously, they must both have taken the same long or short path. This indistinguishability leads, as always in quantum physics, to interference between the long–long and short–short paths. Continuously scanning the phase in one interferometer (at Jussy), while keeping the other one stable, produces a sinusoidal oscillation of the correlation between the photon detections at Satigny and Jussy (Fig. 3). The phases were controlled by the temperature of the fibre-based interferometers.



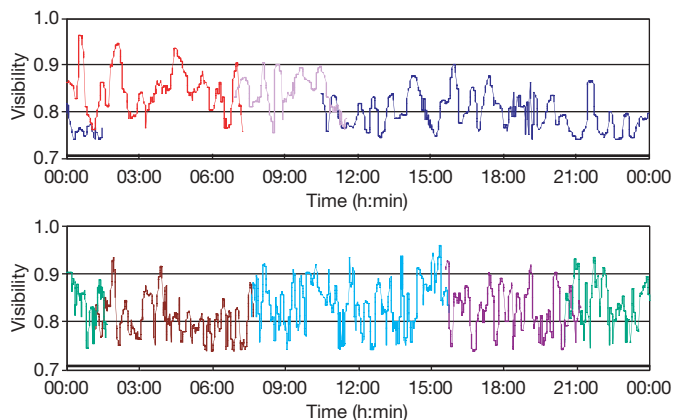
**Figure 2 | Experimental setup.** The source sends pairs of photons from Geneva to two receiving stations through the Swisscom fibre-optic network. The stations are situated in two villages, Satigny and Jussy, that are respectively 8.2 and 10.7 km from Geneva. The direct distance between them is 18.0 km. At each receiving station, the photons pass through identically unbalanced Michelson interferometers and are detected by a single-photon InGaAs avalanche photodiode (APD) (id201, id Quantique). The length of the fibre going to Jussy is 17.5 km. The fibre going to Satigny is only 13.4 km long, so we added a fibre coil of 4.1 km (represented as a loop) to equalize the lengths of the fibres. Having fibres with the same length allows us to satisfy the condition of good alignment ( $\rho \ll 1$ ).  $\delta$  indicates the scanning of the phase of the interferometer at Jussy.



**Figure 3 | Interference fringes.** Interference fringes with a period  $T = 900$  s obtained during a 4-hour measurement, fitted with a sinusoidal function yielding a visibility of  $87.6\% \pm 1.1\%$ . If we subtract the accidental coincidences, the (net) visibility climbs to  $94.1\% \pm 1.0\%$ . This result is significant because the period of the interference fringes remains stable for a very long time, which allows us to fit the entire measurement with a continuous fit and obtain a high visibility value.

Interference fringes were recorded in many runs, which usually lasted several hours (up to 15 hours for the longest run (see Fig. 4)). By juxtaposing several of these measurement runs obtained over several weeks, we covered a 24-hour period with interference fringe periods of  $T = 360$  s with visibilities well above the threshold ( $1/\sqrt{2}$ ) set by the Clauser–Horne–Shimony–Holt Bell inequality<sup>19</sup>. The visibility is large enough to exclude any common-cause explanation. The correlations are thus due either to entanglement, as predicted by quantum physics, or to some hypothetical spooky action at a distance whose speed we wish to bound from below. Because long measurements of fringes with short fringe periods  $T$  are difficult to fit continuously, we fitted the data over a time window corresponding to one-and-a-half fringes and scanned this time window, as explained in the Supplementary information.

The violation of the Bell inequality at all times of day makes it possible to calculate the lower bound for the speed of quantum information in any reference frame. This bound depends on the



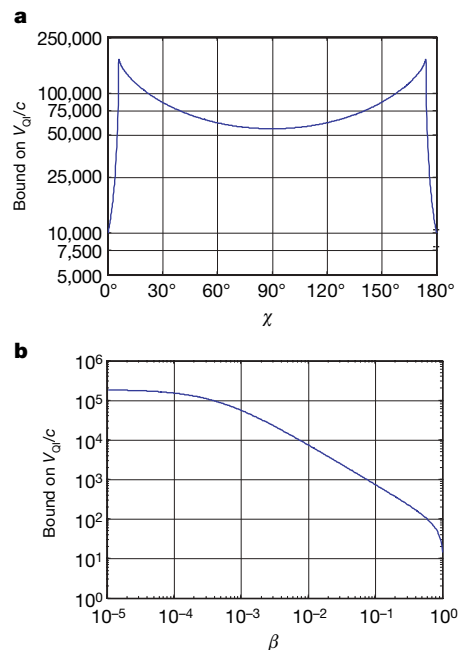
**Figure 4 | Visibility fits.** Visibility fits for several uninterrupted runs obtained at different times of the day. Together these runs cover each moment of the day at least twice. The limitations on the lengths of these measurements were due to the end of the cooling ramp and small instabilities in the set-up that produced short interruptions in the scan. Visibility values remain above the threshold value of  $1/\sqrt{2}$  (black line) set by the Clauser–Horne–Shimony–Holt Bell inequality at all times.

precision of the alignment in the actual experiment; see equation (2). We wished to have a good alignment ( $\rho \ll 1$ ), so the difference in the arrival times of the single photons  $t_{AB}$  was minimized.

First, the length of each fibre between the source and the single-photon detectors was measured. The long fibres (several kilometres in length) were measured using a single-photon optical time-domain reflectometer<sup>20</sup> and the short fibres (less than 500 m in length) were measured using an optical frequency-domain reflectometer<sup>21</sup>. The fibre on the Sagny side was found to be shorter than the other by 4.1 km. We added a fibre coil to the short side (represented as a loop in Fig. 2), reducing this difference to below 1 cm with an uncertainty of 1 cm, which corresponds to a light travel time of 49 ps. To remove any doubt about where exactly the measurement took place, we adjusted the lengths of the fibres from the source to the fibre couplers inside each interferometer and also to the photodiodes (where the photons are detected). Hence, the configuration was totally symmetric.

Next, we considered the chromatic dispersion in the fibres. Chromatic dispersion added an uncertainty in the arrival times, and because the entangled photons were anticorrelated in energy, their time delays were always opposite to each other, which always increased this uncertainty. Chromatic dispersion was measured to be  $18.2 \text{ ps nm}^{-1} \text{ km}^{-1}$  using a chromatic dispersion analyser<sup>22</sup>. For a spectral half-width of  $\Delta\lambda = 0.5 \text{ nm}$  and twice the distance of 17.5 km, this is equivalent to a 319-ps uncertainty. Thus, the overall uncertainty in the relative lengths of the fibres was, when expressed in light travel time,  $t_{AB} = 323 \text{ ps}$ . This, together with the direct distance between the receivers,  $r_{AB} = 18.0 \text{ km}$ , allowed us to estimate the precision of our alignment:  $|\rho| \leq \bar{\rho} = 5.4 \times 10^{-6} \ll 1$ .

Last, we used equation (2) to calculate a lower bound for  $V_{QI}$ . We used the value of  $\bar{\rho}$  just calculated, the period of time  $T = 360$  s needed to observe a Bell violation (corresponding to the interference fringe period), and the angle formed by the axis between the two receiving stations (the A–B axis) and the equatorial ( $x$ – $y$ ) plane,  $\alpha = 5.8^\circ$ . The results are shown in Fig. 5, for certain hypothetically privileged frames. In Fig. 5a, we scan all possible directions  $\chi$ , but set the Earth's relative



**Figure 5 | Lower bounds for the speed of quantum information.** **a**, Bound obtained for  $V_{QI}/c$  as a function of the angle  $\chi$ , for  $\beta = 10^{-3}$ . For  $\chi \lesssim \alpha$  or  $\chi \gtrsim 180^\circ - \alpha$ , the bound is obtained by considering the second case discussed in the text, whereas for  $\alpha \lesssim \chi \lesssim 180^\circ - \alpha$ , the bound is obtained by considering the first case. The bound at  $\chi = 90^\circ$  is  $V_{QI} \geq 54,000c$ . **b**, Bound obtained for  $V_{QI}/c$  as a function of the speed  $\beta$ , for  $\chi = 90^\circ$ . As  $\beta$  tends to zero, our bound on  $V_{QI}/c$  tends to  $1/\bar{\rho}$ .

speed at  $\beta = 10^{-3}$ . A lower bound for  $V_{\text{QI}}$  greater than 10,000 times the speed of light is found in any such reference frame. The imperfect east–west orientation ( $\alpha \neq 0$ ) is responsible for the minimum values of the bound at angles  $\chi$  near  $0^\circ$  and  $180^\circ$ . For smaller Earth speeds, the bound on  $V_{\text{QI}}$  is even larger. However, if  $\beta$  is very large, then the corresponding bound on  $V_{\text{QI}}$  is less stringent. To illustrate this, in Fig. 5b we set  $\chi = 90^\circ$ , that is, with  $\mathbf{v}$  in the equatorial plane, and scan the velocity  $\beta$ . Indeed, for  $\beta \approx 1$ , the bound drops rapidly. Recall, however, that for large values of  $\beta$  we could in principle optimize the alignment  $\rho$  in the experiment to get a better bound on  $V_{\text{QI}}$ . For small values of  $\beta$ , our bound is limited by the inverse of our precision of alignment  $\bar{\rho}$ .

In conclusion, we performed a Bell experiment using entangled photons between two villages separated by 18 km and approximately east–west oriented, with the source located precisely in the middle. The rotation of the Earth allowed us to test all possible hypothetically privileged frames over a period of 24 hours. Two-photon interference fringes with visibilities well above the threshold set by the Bell inequality were observed at all times of day. From these observations we conclude that the nonlocal correlations observed here and in previous experiments<sup>1</sup> are indeed truly nonlocal. To maintain an explanation based on spooky action at a distance we would have to assume that the spooky action propagates at speeds even greater than the bounds obtained in our experiment.

Received 2 April; accepted 30 May 2008.

- Aspect, A. Bell's inequality test: more ideal than ever. *Nature* **398**, 189–190 (1999).
- Aspect, A. *et al.* Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A new violation of Bell's inequalities. *Phys. Rev. Lett.* **49**, 91–94 (1982).
- Tittel, W., Brendel, J., Zbinden, H. & Gisin, N. Violation of Bell inequalities by photons more than 10 km apart. *Phys. Rev. Lett.* **81**, 3563–3566 (1998).
- Weihs, G., Jennewein, T., Simon, C., Weinfurter, H. & Zeilinger, A. Violation of Bell's inequality under strict Einstein locality conditions. *Phys. Rev. Lett.* **81**, 5039–5043 (1998).
- Rowe, M. A. *et al.* Experimental violation of a Bell's inequality with efficient detection. *Nature* **409**, 791–794 (2001).
- Matsukevich, D. N. *et al.* Bell inequality violation with two remote atomic qubits. *Phys. Rev. Lett.* **100**, 150404 (2008).
- Eberhard, P. H. *Quantum Theory and Pictures of Reality* (ed. Schommers, W.) 169–216 (Springer, Berlin, 1989).
- Scarani, V. *et al.* The speed of quantum information and the preferred frame: analysis of experimental data. *Phys. Lett. A* **276**, 1–7 (2000).
- Feinberg, G. Possibility of faster-than-light particles. *Phys. Rev.* **159**, 1089–1105 (1967).
- Bohm, D. A suggested interpretation of the quantum theory in terms of “hidden” variables I. *Phys. Rev.* **85**, 166–179 (1952).
- Bohm, D. A suggested interpretation of the quantum theory in terms of “hidden” variables II. *Phys. Rev.* **85**, 180–193 (1952).
- Bohm, D. & Hiley, B. J. *The Undivided Universe* 293 (Routledge, London, 1993).
- Gisin, N., Scarani, V., Tittel, W. & Zbinden, H. Optical tests of quantum nonlocality: from EPR-Bell tests towards experiments with moving observers. *Ann. Phys.* **9**, 831–841 (2000).
- Zbinden, H. *et al.* Experimental test of nonlocal quantum correlation in relativistic configurations. *Phys. Rev. A* **63**, 022111 (2001).
- Garisto, R. What is the speed of quantum information? Preprint at (<http://arxiv.org/abs/quant-ph/0212078>) (2002).
- Franson, J. D. Bell inequality for position and time. *Phys. Rev. Lett.* **62**, 2205–2208 (1989).
- Tanzilli, S. *et al.* PPLN waveguide for quantum communication. *Eur. Phys. J. D* **18**, 155–160 (2002).
- Thew, R. *et al.* Experimental investigation of the robustness of partially entangled qubits over 11 km. *Phys. Rev. A* **66**, 062304 (2002).
- Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.* **23**, 880–884 (1969).
- Scholder, F., Gautier, J.-D., Wegmüller, M. & Gisin, N. Long-distance OTDR using photon counting and large detection gates at telecom wavelength. *Opt. Commun.* **213**, 57–61 (2002).
- Passy, R. *et al.* Experimental and theoretical investigations of coherent OFDR with semiconductor laser sources. *J. Lightwave Technol.* **12**, 1622–1630 (1994).
- Brendel, J., Gisin, N. & Zbinden, H. in *Proc. 5th Opt. Fiber Measurement Conf.* (eds Boissier, Ch. & Tanguy, E.) 12–17 (Université de Nantes, Nantes, 1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge technical support by J.-D. Gautier and C. Barreiro. The access to the telecommunications network was provided by Swisscom. This work was supported by the Swiss NCCR Quantum Photonics and the European Union project QAP. The image of the Earth in Fig. 1 is a NASA Goddard Space Flight Center Image by R. Stöckli.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to D.S. ([daniel.salarat@physics.unige.ch](mailto:daniel.salarat@physics.unige.ch)).

# Fullerenes from aromatic precursors by surface-catalysed cyclodehydrogenation

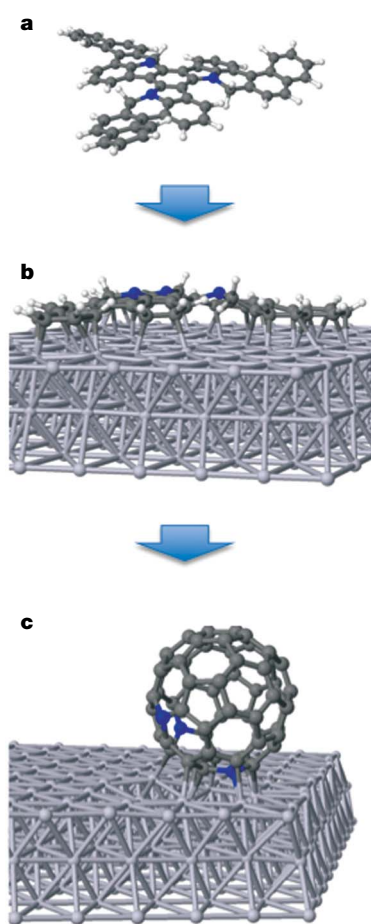
Gonzalo Otero<sup>1\*</sup>, Giulio Biddau<sup>2\*</sup>, Carlos Sánchez-Sánchez<sup>1</sup>, Renaud Caillard<sup>1</sup>, María F. López<sup>1</sup>, Celia Rogero<sup>3</sup>, F. Javier Palomares<sup>1</sup>, Noemí Cabello<sup>4</sup>, Miguel A. Basanta<sup>2</sup>, José Ortega<sup>2</sup>, Javier Méndez<sup>1</sup>, Antonio M. Echavarren<sup>4</sup>, Rubén Pérez<sup>2</sup>, Berta Gómez-Lor<sup>1</sup> & José A. Martín-Gago<sup>1,3</sup>

Graphite vaporization provides an uncontrolled yet efficient means of producing fullerene molecules. However, some fullerene derivatives or unusual fullerene species might only be accessible through rational and controlled synthesis methods. Recently, such an approach has been used<sup>1</sup> to produce isolable amounts of the fullerene C<sub>60</sub> from commercially available starting materials. But the overall process required 11 steps to generate a suitable polycyclic aromatic precursor molecule, which was then cyclodehydrogenated in the gas phase with a yield of only about one per cent. Here we report the formation of C<sub>60</sub> and the triazafullerene C<sub>57</sub>N<sub>3</sub> from aromatic precursors using a highly efficient surface-catalysed cyclodehydrogenation process. We find that after deposition onto a platinum (111) surface and heating to 750 K, the precursors are transformed into the corresponding fullerene and triazafullerene molecules with about 100 per cent yield. We expect that this approach will allow the production of a range of other fullerenes and heterofullerenes<sup>2,3</sup>, once suitable precursors are available. Also, if the process is carried out in an atmosphere containing guest species, it might even allow the encapsulation of atoms or small molecules to form endohedral fullerenes<sup>4,5</sup>.

The surface-catalysed cyclodehydrogenation process we use to transform complex organic polyaromatic precursors of suitable topology into targeted fullerene species is sketched in Fig. 1 (see also Supplementary Fig. 1). Through vacuum thermal evaporation, we first deposit the precursor (C<sub>57</sub>H<sub>33</sub>N<sub>3</sub> (**1**) in the case of triazafullerene and C<sub>60</sub>H<sub>30</sub> (**2**) in the case of fullerene) on a catalytically active metal surface. Subsequent annealing of the sample at 750 K induces a surface reaction that produces the corresponding closed molecule.

The syntheses of the planar precursors used in this work follow a previously published methodology<sup>6–8</sup> (see also Supplementary Information, section 1.1). Their molecular structures are optimized using two different *ab initio* total-energy methods<sup>9,10</sup> (Supplementary Information, Section 1.2); Fig. 1a shows a ball-and-stick model of the optimized geometry we obtain for **1**. Compounds **1** and **2** are markedly twisted, with each of the three lobes adopting a helical shape because steric congestion of their bay positions forces the outer aromatic ring of each lobe to flip up with respect to the plane of the central ring. The twisted conformation can be described by the distance the most external carbon atom is placed above the plane of the central aromatic ring ( $\Delta z$ ), and by the twisting angle of each molecular branch. These values are respectively found to be 0.08 nm and 20.5° for **1** and 0.09 nm and 30.5° for **2**.

We have probed the cyclodehydrogenation process using scanning tunnelling microscopy (STM), X-ray photoemission spectroscopy



**Figure 1 | Optimized geometrical structure of the C<sub>57</sub>H<sub>33</sub>N<sub>3</sub> molecule at the different stages of the process.** **a**, Representation of the optimized molecular structure of the C<sub>57</sub>H<sub>33</sub>N<sub>3</sub> molecule. The wings of the molecule display a noticeable helicoidal twist. **b**, The same molecule adsorbed at room temperature (300 K) on the platinum (111) surface. The final relaxed configuration strongly depends on the adsorption site. The structure shown corresponds to one of the most energetically favourable sites, where the interaction with the metal substrate removes the twist and favours a markedly planar configuration for the molecule. **c**, The optimized structure for the C<sub>57</sub>N<sub>3</sub> triazafullerene formed after the cyclodehydrogenation process at 750 K. Blue balls represent the nitrogen atoms in the structure.

<sup>1</sup>Instituto de Ciencia de Materiales de Madrid (CSIC), Cantoblanco, 28049 Madrid, Spain. <sup>2</sup>Departamento de Física Teórica de la Materia Condensada, Universidad Autónoma de Madrid, 28049 Madrid, Spain. <sup>3</sup>Centro de Astrobiología (CSIC-INTA), Carretera de Torrejón a Ajalvir, km 4, 28850 Torrejón de Ardoz, Madrid, Spain. <sup>4</sup>Institute of Chemical Research of Catalonia (ICIQ), Avinguda Paisos Catalans 16, 43007 Tarragona, Spain.

\*These authors contributed equally to this work.

(XPS) and thermal programmed desorption (TPD) in an ultrahigh vacuum (UHV) environment. Figure 2a shows an STM image of **1** deposited on a platinum (111) surface with the ball-and-stick model of the adsorbed molecule superimposed. The size and topology of the STM image, with three wings characterizing the molecular shape, confirm that the molecules do not fragment, but retain their planar structure upon adsorption. A more intense protrusion (lobe) is imaged in each of the wings. These lobes are not centred; rather, their positions correspond to those of the hexagonal external rings, as can be seen by comparing the superimposed molecule with the STM image in Fig. 2a.

In a first approximation, we might speculate that the lobes correspond to the molecular orbitals of the free molecule. However, the detailed relaxed geometry is far from that of the isolated molecule: our *ab initio* calculations on a large  $10 \times 10$  unit cell show that the interaction with the platinum surface is strong enough to deform the structure of the molecule. The competition between the  $\pi$  bonding of the rings with the surface and the energetic costs of the sterically induced molecular deformations results in a very complicated energy landscape, with barriers between local minima determined by the adsorption geometry. The general trend is a reduction of the molecular buckling (for example,  $\Delta z$  is reduced from 0.08 nm to 0.02 nm for **1**) as the molecular wings approach the surface to maximize the number of bonds with the surface (see also the relaxed configuration shown in Fig. 1b). The effect brings the hydrogen atoms on the wings closer to the surface, and thus plays an important role in the temperature-induced fullerene formation by surface-catalysed dehydrogenation. The strong molecule–surface interaction suggested by the calculations can also explain why we fail to observe long-range ordering of adsorbed molecules or preferential adsorption at step edges (which would require mobility of the molecule on the platinum surface). Similarly, the observation that in some imaged molecules

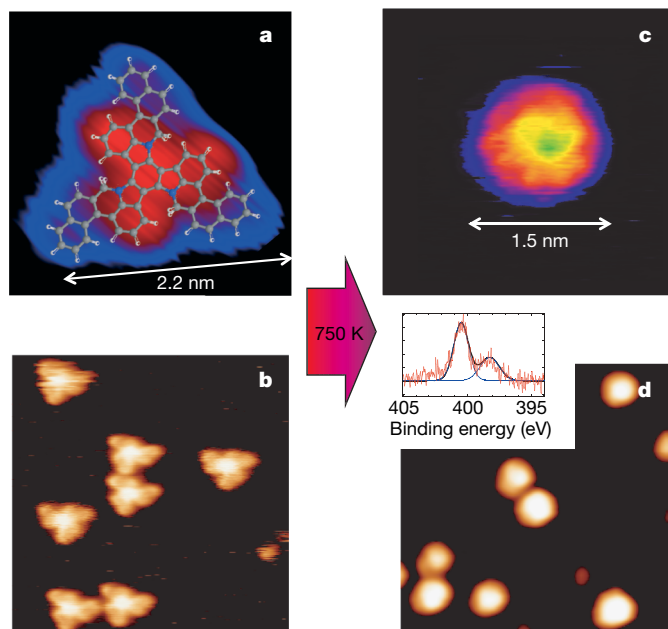
the different lobes have different apparent heights is consistent with our calculations, which indicate that the final molecular geometry strongly depends on the adsorption site.

As deposited on the platinum (111) surface (Fig. 2a), the heteropolyporene **1** (and similarly the polyporene **2**) presents a total apparent height in the STM images of about 0.14 nm, with no significant dependence on bias voltage. The triangular base length of the molecule is about 2.2 nm. Upon annealing the sample at 750 K, a surface reaction takes place and the triangular molecules imaged on the surface transform into round molecules with an apparent height of about 0.38 nm and a diameter of 1.5 nm (Fig. 2c). Evaporation of commercial  $C_{60}$  on a platinum surface and imaging under the same experimental conditions results in very similar STM images for the adsorbed fullerenes (Supplementary Fig. 2). A height of about 0.4 nm and a diameter of about 1.35 nm have been reported for  $C_{60}$  deposited on a platinum (110) surface<sup>11</sup>, and a height of about 0.31 nm for  $C_{60}$  deposited on a palladium (110) surface<sup>12</sup>. In both these cases, the low height appears after surface annealing and suggests strong bonding of the molecule to the surface, in agreement with suggestions that the interaction of a fullerene with a platinum surface is strongly covalent and that the molecule tries to maximize the number of surface bonds<sup>13</sup>. We also note that the molecular orbitals visible in the STM images are similar to those reported for  $C_{60}$  on platinum<sup>14</sup>, silicon<sup>15</sup> and gold<sup>16</sup> surfaces and are fairly well reproduced by our calculations (Supplementary Information, section 2.2); they might thus be considered as an indicator of fullerene formation.

The cyclization process shows efficiency close to 100%; that is, all the adsorbed molecules are transformed into fullerenes (Supplementary Information, section 2.3). This is illustrated by the images in Figs 2b, d of the surface covered with about 0.2 monolayers of precursor before and after annealing, respectively: the molecular coverage does not change as a result of the temperature-induced reaction, with all of the triangular molecules converted into round fullerenes.

Further evidence for the thermally induced cyclization of **1** is provided by XPS spectra of the surface covered with about 0.8 monolayers of precursor (Fig. 2d inset). Even though this corresponds to a nitrogen coverage of only about 0.04 monolayers with respect to the platinum surface, weak emission from the nitrogen 1s core level is detected and confirms that after annealing, nitrogen is present on the surface (and is presumably associated with the unique topographic STM surface features, the fullerene molecules). Information on the chemical state of the nitrogen atoms is obtained from a core-level line-shape analysis of the nitrogen 1s spectrum, with its two peaks. The binding energy of the main peak at 400.6 eV corresponds to substitutional nitrogen in a graphite sheet, as reported for nitrogen bonded to carbon atoms in  $sp^2$  hybridization<sup>17</sup>. This is the nitrogen coordination in the triazafullerene we have formed. The binding energy of the small peak at 398.2 eV is related to  $PtN_x$  compounds, in which nitrogen is chemisorbed<sup>18</sup>. Therefore, the smaller peak can be attributed to nitrogen atoms in the fullerene cage interacting with the platinum surface. This core-level analysis thus indicates that about one-third of the nitrogen atoms of the triazafullerene  $C_{57}N_3$  interact with the surface, whereas, as a consequence of the heterofullerene curvature (see Fig. 1c), the rest are not in contact with the surface. The XPS observations also confirm the 100% efficiency of the cyclodehydrogenation process because the nitrogen 1s intensity does not change significantly after annealing, indicating that all the deposited molecules have cyclized (Supplementary Information, section 2.5).

The strong covalent interaction between the precursor molecules and the platinum surface appears important for the efficiency of the process. To test this idea, we evaporate **1** and **2** on a gold (111) surface that is quite inert and does not promote the cyclization reaction. The molecule–substrate interaction is negligible on this surface and, consequently, molecular diffusion at room temperature is enhanced, inhibiting the dehydrogenation process. The cyclization efficiency is reduced in this case to about 1%, with most of the molecules



**Figure 2 | STM images of the cyclization process.** **a**, The  $C_{57}H_{33}N_3$  molecule deposited on the platinum (111) surface, with the atomic structure superimposed following the calculations. **b**,  $10 \times 10 \text{ nm}^2$  image of the platinum (111) surface after evaporation of about 0.2 monolayers of  $C_{57}H_{33}N_3$  molecules. **c**, An isolated  $C_{57}N_3$  molecule obtained after annealing the surface at 750 K. The cyclization process is inferred from the appearance of different molecular orbitals and changes in the shape and size of the imaged molecule. **d**,  $10 \times 10 \text{ nm}^2$  image of the surface shown in **b**, but after annealing at 750 K. Inset, XPS spectrum of about 0.8 monolayers of  $C_{57}H_{33}N_3$  molecules after annealing at 750 K. The best-fit curve of the spectrum, together with the different components used in the fit, is also shown.

decomposing. In the test with polyarene **2**, the few isolated  $C_{60}$  molecules that are detected on the gold surface are found to have molecular orbitals similar to those previously reported<sup>15</sup>.

The thermally induced surface reaction generating fullerenes consists of a series of dehydrogenation reactions. Recently, dehydrogenation of adsorbed aromatic molecules was attributed to activation by the electronic current from the tip of a scanning tunnelling microscope<sup>19</sup>. To exclude such an activation mechanism and to confirm a surface-catalysed dehydrogenation process in our system, we measure the dihydrogen signal in a mass spectrometer as a function of annealing temperature. To avoid experimental artefacts due to dihydrogen desorption from the experimental apparatus, we synthesize 2,3,7,8,12,13-hexadeutero-10,15-dihydro-5H-diindolo[3,2-*a*:3',2'-*c*]carbazole ( $C_{57}H_{27}D_6N_3$ , or **1-d<sub>6</sub>**), a deuterated analogue of  $C_{57}H_{33}N_3$  (**1**) incorporating six deuterium atoms at strategic positions. Figure 3 shows the evolution of  $m/z$  3 (HD) and  $m/z$  4 ( $D_2$ ) mass signals as a function of temperature. The baseline corresponds to the  $m/z$  3 and 4 mass signals recorded when annealing the clean platinum surface (that is, before any molecular deposition) up to 750 K. The mass spectrum shows that a rapid desorption process takes place at about 700 K and should be associated with the molecular transformation shown in Fig. 2. The STM observations agree with the mass spectrometry results (Fig. 3), which indicate that there is neither modification of the molecular shape nor dehydrogenation below 400 K and that we could have partial dehydrogenation at about 500 K. STM images recorded after annealing between 400 and 600 K show asymmetrical molecules that could correspond to intermediates, that is, partially closed fullerenes that could be regarded as open-cage fullerenes<sup>20,21</sup>.

The various experimental observations presented above provide insight into the cyclization mechanism. The observed formation of HD ( $m/z$  3) can be attributed to the exchange of  $H_2$  and  $D_2$ , which is known to be catalysed by platinum with an activation energy of 0.23 eV (ref. 22) and is enhanced in our system owing to the relatively high temperature of the dehydrogenation process. Therefore, hydrogen and deuterium are gathered by the platinum surface, where they diffuse, react with each other and finally desorb as  $H_2$ , HD and  $D_2$ . The platinum-induced dehydrogenation of the precursor that produces

the hydrogen and deuterium atoms is facilitated by the strong interaction of the molecule with the platinum surface. According to our density functional theory calculations, this interaction distorts the molecule in such a way that the molecular wings approach the surface and enable platinum-catalysed dissociation of hydrogen atoms from the adsorbed molecule (Fig. 1b). Conversely, if the molecule does not interact strongly with the surface, as in the case of a gold (111) surface, cyclodehydrogenation does not take place. After consecutive release of the hydrogen atoms to the surface, the molecule is assumed to spontaneously cyclize. We note that although a complete density functional theory simulation of this complex process is not possible, our theoretical calculations for the  $C_{60}H_{30}$  free molecule provides evidence that dehydrogenation induces a strong tendency towards cyclization. The removal of hydrogen atoms close to the central ring triggers conformational changes such that the molecular wings approach each other and a fullerene semi-cage is formed in a process that proceeds without an energy barrier. Furthermore, a thorough study of the completely dehydrogenated molecule shows that the different possible paths leading from the planar conformation to the complete fullerene cage have energy barriers less than 0.3 eV, which can be easily overcome with the thermal energy available during annealing (Supplementary Information, section 2.4).

To explore the feasibility of both our controlled synthesis method outside an UHV environment and the potential for scale-up, we use platinum nanopowder instead of a single-crystal substrate. In this experiment, compound **2** is deposited from a  $CHCl_3$  solution onto a large excess of platinum nanopowder and heated under vacuum ( $10^{-4}$  mbar) at 700 K for 20 min. For this sample, the laser desorption–ionization time-of-flight mass spectrum in negative-ion mode clearly shows the presence of  $C_{60}$  (Supplementary Information, section 2.6).

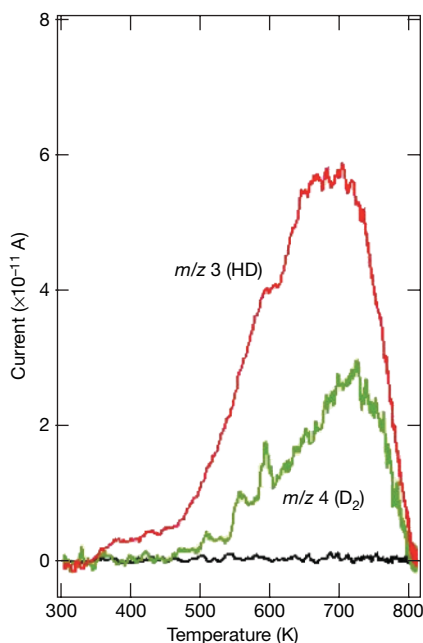
## METHODS SUMMARY

The syntheses of  $C_{60}H_{30}$  and  $C_{57}H_{33}N_3$  were performed by alkylation of 10,15-dihydro-5H-diindeno[1,2-*a*:1',2'-*c*]fluorene (truxene) and, respectively, 10,15-dihydro-5H-diindolo[3,2-*a*:3',2'-*c*]carbazole (5,10,15-triazatruxene)<sup>6–8</sup>. Cyclization experiments have been performed *in situ* in an UHV system with a base pressure of  $1 \times 10^{-10}$  mbar equipped with Auger electron spectroscopy, low-energy electron diffraction optics, STM at room temperature (300 K) and TPD. The molecules were deposited *in situ* by thermal evaporation on a previously atomically cleaned platinum (111) surface. Calculations are based on density functional theory as implemented in two different total-energy pseudopotential codes: FIREBALL (ref. 9), where a local orbital basis is used to expand the electronic wavefunctions, and the plane-wave-basis approach CASTEP (ref. 10). Laser desorption–ionization time-of-flight experiments were performed using a Bruker Daltonics Autoflex mass spectrometer equipped with a nitrogen laser (337 nm). The instrument was operated in negative-ion, reflectron mode.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 13 July 2007; accepted 19 June 2008.

1. Scott, L. T. *et al.* A rational chemical synthesis of  $C_{60}$ . *Science* **295**, 1500–1503 (2002).
2. Hummelen, J. C., Knight, B., Pavlovich, J., González, R. & Wudl, F. Isolation of the heterofullerene  $C_{59}N$  as its dimer ( $C_{59}N$ )<sub>2</sub>. *Science* **269**, 1554–1556 (1995).
3. Vostrowsky, O. & Hirsch, A. Heterofullerenes. *Chem. Rev.* **106**, 5191–5207 (2006).
4. Komatsu, K., Murata, M. & Murata, Y. Encapsulation of molecular hydrogen in fullerene  $C_{60}$  by organic synthesis. *Science* **307**, 238–240 (2005).
5. Iwamatsu, S.-I. *et al.* Carbon monoxide inside an open-cage fullerene. *Angew. Chem. Int. Ed.* **45**, 5337–5340 (2006).
6. Gómez-Lor, B., de Frutos, Ó. & Echavarren, A. M. Synthesis of 'crushed fullerene'  $C_{60}H_{30}$ . *Chem. Commun.* 2431–2432 (1999).
7. Gómez-Lor, B. *et al.* Zipping up 'the crushed fullerene'  $C_{60}H_{30}$ :  $C_{60}$  by fifteen-fold, consecutive intramolecular  $H_2$  losses. *Chem. Commun.* 370–371 (2002).
8. Gómez-Lor, B. & Echavarren, A. M. Synthesis of a triaza analogue of crushed fullerene by intramolecular palladium catalysed arylation. *Org. Lett.* **6**, 2993–2996 (2004).
9. Jelinek, P., Wang, H., Lewis, J. P., Sankey, O. F. & Ortega, J. Multicenter approach to the exchange-correlation interactions in ab initio tight-binding methods. *Phys. Rev. B* **71**, 235101 (2005).



**Figure 3 | TPD experiment of the cyclodehydrogenation process.** Temperature evolution of the masses corresponding to HD ( $m/z$  3) and  $D_2$  ( $m/z$  4), recorded after depositing  $C_{57}H_{27}D_6N_3$  on the surface. Thermal desorption takes place at around 500 K. The  $m/z$  4 signal has been scaled by a factor of eight. The black line is the result of recording the HD signal without depositing molecules on the platinum surface.

10. Segall, M. D. *et al.* First-principles simulation: ideas, illustrations and the CASTEP code. *J. Phys. Condens. Matter* **14**, 2717–2744 (2002).
11. Orzali, T., Petukhov, M., Sami, M. & Tondello, E. STM study of the initial stages of C<sub>60</sub> adsorption on the Pt(110)-(1×2) surface. *Appl. Surf. Sci.* **252**, 5534–5537 (2006).
12. Weckesser, J., Barth, J. V. & Kern, K. Mobility and bonding transition of C<sub>60</sub> on Pd(110). *Phys. Rev. B* **64**, 161403 (2001).
13. Felici, R. *et al.* X-ray-diffraction characterization of Pt(111) surface nanopatterning induced by C<sub>60</sub> adsorption. *Nature Mater.* **4**, 688–692 (2005).
14. Orzali, T. *et al.* Temperature-dependent self-assemblies of C<sub>60</sub> on (1×2)-Pt(110): A STM/DFT investigation. *J. Phys. Chem. C* **112**, 378–390 (2008).
15. Pascual, J. I. *et al.* Seeing molecular orbitals. *Chem. Phys. Lett.* **321**, 78–82 (2000).
16. Schull, G. & Berndt, R. Orientationally ordered (7×7) superstructure of C<sub>60</sub> on Au(111). *Phys. Rev. Lett.* **99**, 226105 (2007).
17. Hellgren, N. *et al.* Electronic structure of carbon nitride thin films studied by X-ray spectroscopy techniques. *Thin Solid Films* **471**, 19–34 (2005).
18. Soto, G. Synthesis of PtN<sub>x</sub> films by reactive laser ablation. *Mater. Lett.* **58**, 2178–2180 (2004).
19. Lesnard, H., Bocquet, M. L. & Lorente, N. Dehydrogenation of aromatic molecules under a scanning tunnelling microscope: Pathways and inelastic spectroscopy simulations. *J. Am. Chem. Soc.* **129**, 4298–4305 (2007).
20. Rubin, Y. Ring opening reactions of fullerenes: Designed approaches to endohedral metal complexes. *Top. Curr. Chem.* **199**, 67–91 (1999).
21. Murata, Y., Murata, M. & Komatsu, K. Synthesis, structure, and properties of novel open-cage fullerenes having heteroatom(s) on the rim of the orifice. *Chem. Eur. J.* **9**, 1600–1609 (2003).
22. Gale, R. J., Salmeron, M. & Somorjai, G. A. Variation of surface-reaction probability with reactant angle of incidence: molecular-beam study of asymmetry of stepped platinum crystal-surfaces for H-H bond breaking. *Phys. Rev. Lett.* **38**, 1027–1029 (1977).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Authors acknowledge financial support from the national Spanish funding agency DGICYT-MEC (programmes MAT, CONSOLIDER and CTQ) and the ICIQ Foundation. Computer time was provided by the Spanish National Supercomputing Network at the MareNostrum (BSC) and Magerit (CESVIMA) supercomputers.

**Author Contributions** G.O. performed the experimental work related to the UHV system. G.B. performed the *ab initio* calculations. C.S.-S., R.C., M.F.L., C.R., F.J.P. and J.M. helped with STM-STS, TDS and XPS experiments and preliminary synchrotron radiation measurements. A.M.E. and B.G.-L. designed and synthesized the planar precursors of (hetero)fullerenes. N.C. performed the mass spectrometry experiments. M.A.B. and J.O. wrote part of the calculation code and input and made preliminary calculations. The work was coordinated by B.G.-L. (chemistry), R.P. (theory) and J.A.M.-G. (experiment: UHV-STM molecular deposition, characterization and cyclization).

**Author Information** Reprints and permissions information is available at [npg.nature.com/reprints](http://npg.nature.com/reprints). Correspondence and request of further information should be addressed to B.G.-L. ([bgl@icmm.csic.es](mailto:bgl@icmm.csic.es)) or J.A.M.-G. ([gago@icmm.csic.es](mailto:gago@icmm.csic.es)).

## METHODS

The syntheses of  $C_{60}H_{30}$  and  $C_{57}H_{33}N_3$  were performed by alkylation of 10,15-dihydro-5*H*-diindeno[1,2-*a*:1',2'-*c*]fluorene (truxene) and, respectively, 10,15-dihydro-5*H*-diindolo[3,2-*a*:3',2'-*c*]carbazole (5,10,15-triazatruxene) with 1-bromo-2-bromomethylnaphthalene, followed by a triple palladium-catalysed arylation as previously reported<sup>6–8</sup>. The synthesis of the hexadeuterated analogue  $C_{57}H_{27}D_6N_3$  was performed according to the same procedure but starting from a hexadeutero-5,10,15-triazatruxene, obtained by reductive deuteriobromination of hexabromo-5,10,15-triazatruxene with formic acid- $d_2$  and  $Et_3N$ .

Calculations were based on density functional theory as implemented in two different total-energy pseudopotential codes: FIREBALL<sup>9</sup>, where a local orbital basis is used to expand the electronic wavefunctions and CASTEP<sup>10</sup>, a plane-wave-basis approach. FIREBALL provides the balance of accuracy and efficiency needed to perform calculations with the extremely large unit cells (involving almost 400 atoms) that are required to study the adsorption of the planar fullerenes on the metal substrates. Its reliability has been checked against CASTEP results for the different free molecules.

To model the adsorption of the planar fullerene-precursors on the platinum (111) surface, we considered a  $10 \times 10$  periodic slab including the molecule and three metal layers. Only the  $\Gamma$  point was included in the sampling of the Brillouin zone. The FIREBALL code with the local density approximation for the exchange-correlation functional was used for these calculations. All atoms in the molecule and the two upper metal layers were allowed to relax to their ground-state configuration with convergence criteria for the total energy and forces of  $10^{-4}$  eV and  $0.05 \text{ eV } \text{\AA}^{-1}$ , respectively.

UHV experiments were performed in a system that had a base pressure of  $1 \times 10^{-10}$  mbar, and Auger electron spectroscopy, low-energy electron diffraction, STM at room temperature and TPD techniques were used.

We prepared the clean platinum (111) surface through repeated cycles of  $Ar^+$  bombardment and annealing at up to 1,250 K for 15 min in an  $O_2$  atmosphere ( $1 \times 10^{-5}$  mbar). The last annealing cycle was performed after removing the  $O_2$  atmosphere from the chamber and keeping the total pressure at around  $10^{-10}$  mbar. To prepare the clean gold (111) surface, we used the same methodology but annealed at 700 K without  $O_2$  atmosphere in UHV (pressure during annealing,  $4 \times 10^{-10}$  mbar). To check for surface cleanliness we used Auger electron spectroscopy, low-energy electron diffraction and STM techniques. Typical values for STM images were about +500 mV (on the sample) and 0.2 nA. XPS spectra were recorded in a different system equipped with a SPECS Phoibos150 electron spectrometer and a delay-line detector in the nine-segment mode, using monochromatic Al  $K\alpha$  radiation (1486.74 eV). The pressure in the analysis chamber was less than  $5 \times 10^{-10}$  mbar. High-resolution XPS spectra, such as that shown in Fig. 2, were obtained with a pass energy of 10 eV, which provided an overall energy resolution of 0.39 eV as measured at the Fermi edge of a silver reference sample.

To evaporate the organic molecules we used a homemade tantalum dispenser, which was calibrated with a thermocouple to fix the evaporation temperature accurately at 675 K. The molecular source was previously degassed without molecules up to 1,000 K in high vacuum ( $1 \times 10^{-7}$  mbar). The  $C_{60}H_{30}$ ,  $C_{57}H_{33}N_3$ ,  $C_{57}H_{27}D_6N_3$  and commercial  $C_{60}$  (98% purity; Sigma) molecules were degassed at 660 K in UHV ( $3 \times 10^{-10}$  mbar) for two days to eliminate any organics used in the synthesis.

We performed the synthesis of  $C_{60}$  over platinum nanopowder by dropping a solution of **2** (0.5 mg) in  $CHCl_3$  over 20 mg of activated platinum powder (approximate particle size, 100 nm; Aldrich). The solvent was evaporated under a nitrogen flow and the mixture was heated at 700 K in a Carbolite CTF/12/65/550 tube furnace with a temperature controller device for 20 min under vacuum conditions ( $10^{-4}$  mbar).

# North American ice-sheet dynamics and the onset of 100,000-year glacial cycles

R. Bintanja<sup>1,2</sup> & R. S. W. van de Wal<sup>2</sup>

The onset of major glaciations in the Northern Hemisphere about 2.7 million years ago<sup>1</sup> was most probably induced by climate cooling during the late Pliocene epoch<sup>2,3</sup>. These glaciations, during which the Northern Hemisphere ice sheets successively expanded and retreated, are superimposed on this long-term climate trend, and have been linked to variations in the Earth's orbital parameters<sup>4</sup>. One intriguing problem associated with orbitally driven glacial cycles is the transition from 41,000-year to 100,000-year climatic cycles that occurred without an apparent change in insolation forcing<sup>5</sup>. Several hypotheses have been proposed to explain the transition, both including and excluding ice-sheet dynamics<sup>6–10</sup>. Difficulties in finding a conclusive answer to this palaeoclimatic problem are related to the lack of sufficiently long records of ice-sheet volume or sea level. Here we use a comprehensive ice-sheet model and a simple ocean-temperature model<sup>11</sup> to extract three-million-year mutually consistent records of surface air temperature, ice volume and sea level from marine benthic oxygen isotopes<sup>12</sup>. Although these records and their relative phasings are subject to considerable uncertainty owing to limited availability of palaeoclimate constraints, the results suggest that the gradual emergence of the 100,000-year cycles can be attributed to the increased ability of the merged North American ice sheets to survive insolation maxima and reach continental-scale size. The oversized, wet-based ice sheet probably responded to the subsequent insolation maximum by rapid thinning through increased basal-sliding<sup>13,14</sup>, thereby initiating a glacial termination. Based on our assessment of the temporal changes in air temperature and ice volume during individual glacials, we demonstrate the importance of ice dynamics and ice–climate interactions in establishing the 100,000-year glacial cycles, with enhanced North American ice-sheet growth and the subsequent merging of the ice sheets being key elements.

This onset of major glaciations has been inferred primarily from analyses of marine oxygen isotope ratio ( $\delta^{18}\text{O}$ ) records<sup>15</sup>. The glaciations have been confidently linked to variations in the Earth's orbital parameters—that is, precession of the equinoxes (with periods of 19,000 and 23,000 years; 19 and 23 kyr), axial obliquity (41 kyr) and orbital eccentricity (100 kyr)<sup>4</sup>. However, difficulties associated with the theory of orbitally driven glacial cycles remain. One of the more intriguing problems relates to the observation that marine  $\delta^{18}\text{O}$  varied mainly at obliquity timescales before  $\sim 1$  Myr ago<sup>16</sup> and at 100-kyr timescales thereafter. This transition from 41-kyr to 100-kyr climatic cycles is generally known as the mid-Pleistocene transition (MPT), and occurred without an apparent change in insolation forcing. Numerous explanations have been proposed for this change, such as a nonlinear ice-sheet response to a long-term cooling trend<sup>6</sup>, spatial and temporal changes in ice-sheet substratum<sup>7</sup>, sea-ice effects on ablation and accumulation<sup>8</sup>, orbital-inclination forcing induced

by extraterrestrial dust<sup>9</sup>, and a change in the margin of the Antarctic ice sheet from land-based to marine-based<sup>10</sup>.

The main difficulty in finding conclusive answers to this long-standing palaeoclimatic question is that sufficiently long records of ice-sheet volume or sea level are currently unavailable. Hence, existing hypotheses are mostly based on benthic  $\delta^{18}\text{O}$  records. However, these are influenced not only by the mean oceanic  $\delta^{18}\text{O}$  (governed by Northern Hemisphere ice-sheet storage) but also by local deep-water temperature<sup>17</sup>, and marine  $\delta^{18}\text{O}$  records may therefore be poor indicators of ice-sheet fluctuations in terms of their volume<sup>11</sup> and phase lag owing to the temperature overprint<sup>18–20</sup>.

We have applied a recently developed model-based method<sup>11</sup> to separate these two components, which involves an inverse technique in conjunction with an ice-sheet model coupled to a simple deep-water temperature model (Supplementary Information and Methods). Using the LR04 stack marine  $\delta^{18}\text{O}$  record<sup>12</sup>, we have reconstructed mutually consistent 3-Myr time series of surface air temperature (continental and annual mean between 40° and 80° N), ice-sheet volume and sea level. We have estimated the phasing between them to deduce causes and effects, and infer how the climate variables and phasings have changed over the MPT. These reconstructions are subject to considerable uncertainty<sup>11</sup>, in particular before 0.4 Myr ago, which is due mainly to limited palaeoclimatic constraints (Supplementary Information) and to the simplicity of the ocean-temperature model (and other assumptions<sup>11</sup>).

The reconstructed marine isotope signal increased steadily over the past 3 Myr (Fig. 1a), indicative of both climatic cooling (Fig. 1b) and growing ice sheets (Fig. 1c) (Supplementary Data). Its amplitude (reflecting the contrast between interglacial and glacial states) also increased with time, with noticeable changes occurring around 2.7 Myr ago<sup>21</sup> indicating the onset of Northern Hemisphere glaciation (Fig. 1c), and again around 1 Myr ago (MPT). The inferred overall 3-Myr cooling trend is generally consistent with the gradual cooling of alkenone-derived equatorial sea surface temperature from deep-sea sediment cores<sup>3</sup>, although our mid-latitude/subarctic cooling is stronger, most probably implying an increase in meridional temperature gradient over the 3-Myr period. Our reconstructed deep-sea cooling of about 2–2.5 °C agrees with Mg/Ca deep-sea temperature estimates<sup>2</sup> (Supplementary Information).

According to our calculations, the inception of the Northern Hemisphere ice sheets at about 2.7 Myr ago may be linked to air temperatures dropping below  $-5$  °C (relative to the present), with summer temperatures over inception-sensitive regions being the critical factor<sup>22</sup>. The modelled timing of Northern Hemisphere inception coincides with the first occurrence of ice-rafted debris in the North Atlantic<sup>23</sup>, which indicates calving ice sheets bordering the ocean. The amplitude of the temperature fluctuations hardly increased (Fig. 1b) during this inception period, suggesting that the emerging Northern

<sup>1</sup>Royal Netherlands Meteorological Institute (KNMI), Wilhelminalaan 10, 3732 GK De Bilt, The Netherlands. <sup>2</sup>Institute for Marine and Atmospheric Research Utrecht (IMAU), Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands.

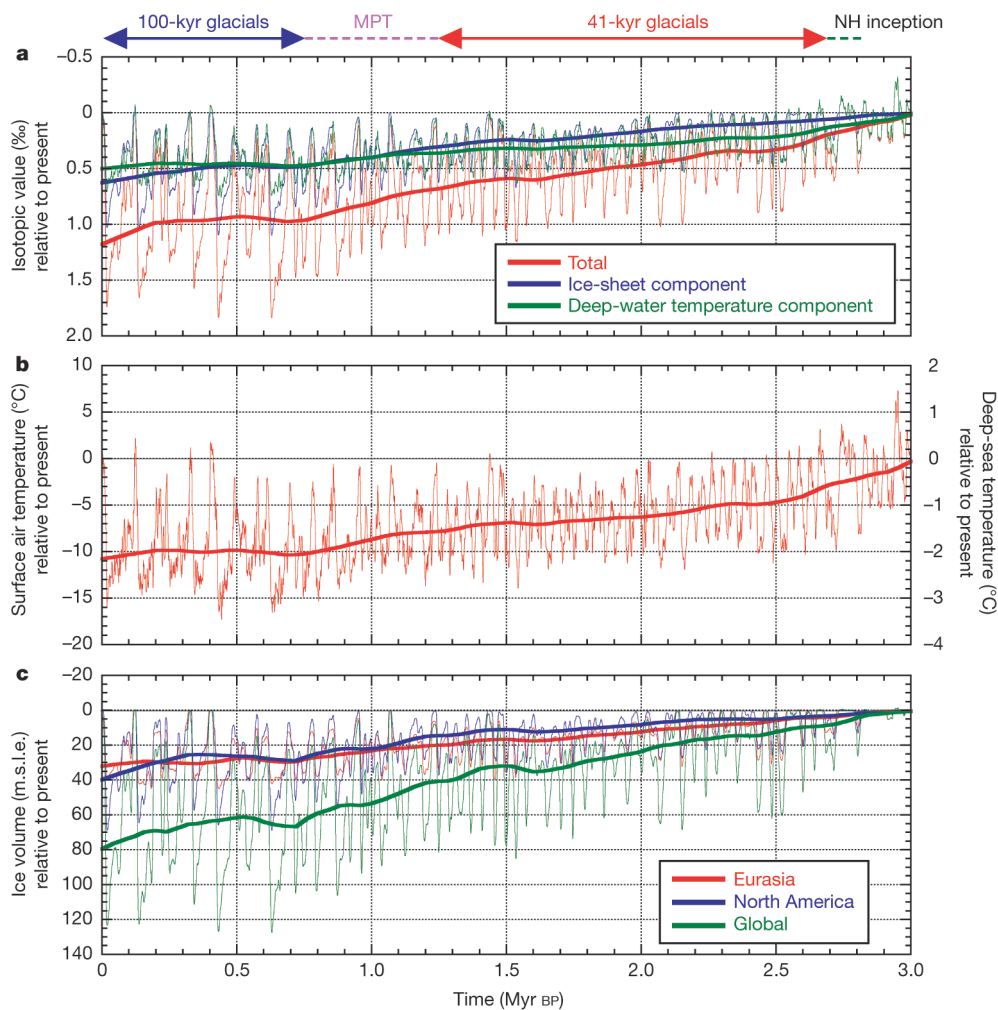
Hemisphere ice sheets accounted for the enhanced  $\delta^{18}\text{O}$  variability around 2.7 Myr ago.

Following Northern Hemisphere glacial inception, a more gradual cooling facilitated steady ice-sheet expansion over the past 2.5 Myr (in the long-term average; thick lines in Fig. 1a, b). Before the MPT, the Eurasian (EAS) ice sheets contributed most to the total Northern Hemisphere ice volume; then, with the appearance of 100-kyr glacials, the North American (NAM) ice sheets took over. According to our model calculations, the EAS and NAM ice sheets on average stored 19.0 and 15.8 m sea level equivalent (m.s.l.e.), respectively, between 1 and 2 Myr ago, and 25.9 and 29.6 m.s.l.e. in the past 1 Myr, suggesting that the MPT somehow involved a change in EAS and NAM ice-sheet characteristics. With the emergence of the 100-kyr cycles during the MPT, the  $\delta^{18}\text{O}$  signal again exhibits a marked increase in amplitude. Our results suggest that this can be attributed mainly to an increase in the glacial–interglacial range of ice volume and global sea level (Fig. 1c), as air temperature fluctuations exhibit only a modest increase in amplitude.

Most theories about the causes and origin of ice ages (glacials), including those involving the MPT, rely heavily on assumptions regarding the phasing of temperature, ice volume and  $\delta^{18}\text{O}$  fluctuations relative to each other and to insolation components<sup>19,21</sup>. Here we estimate these phasings using Blackman–Tukey cross-spectral analysis to determine the phase and coherence spectra relative to

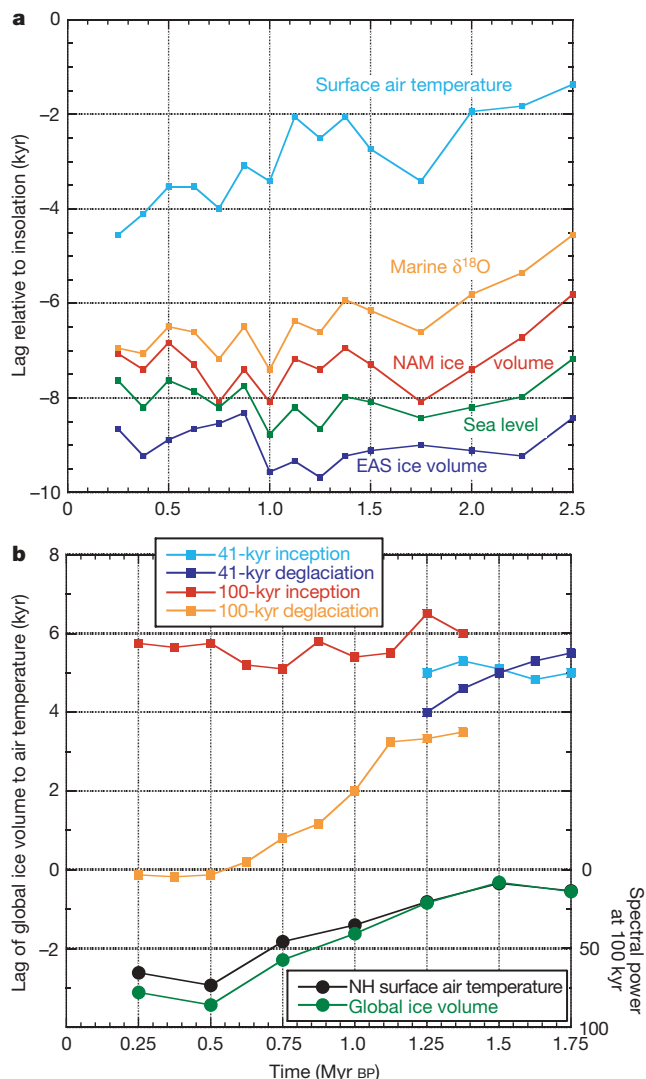
the 21 June insolation at 65° N. The results suggest that temperature exhibited a relatively small lag relative to insolation, which increased from 1.5 kyr at 2.5 Myr ago to 4.5 kyr at present (Fig. 2a), presumably caused by the thermal effects of expanding ice sheets. Because temperature and  $\text{CO}_2$  most probably varied roughly in phase<sup>19</sup>, our results indicate that obliquity forced Northern Hemisphere temperatures—most probably through mechanisms associated with the global carbon cycle—then drove the Northern Hemisphere ice sheets. Global ice volume lagged behind insolation and temperature by several millennia, with the EAS ice sheets exhibiting the slowest response because of colder and drier ambient conditions.

The average lag of ice volume behind surface air temperature decreased from about 6 kyr at 2.5 Myr ago to slightly more than 3 kyr at present (Fig. 2a), with the strongest decrease occurring after the MPT. According to our calculations, the Northern Hemisphere needed to cool about 5 °C before ice-sheet inception occurred. Also, ice sheets cannot expand faster than the rate at which mass is gained through snow accumulation. Both factors contribute to ice volume lagging temperature by 5–7 kyr during inception stages, both for 41-kyr and 100-kyr glaciations, indicating that ice-sheet growth merely followed the temperature signal. The only exception to ice volume lagging air temperature involves 100-kyr deglaciations; the associated lag dropped to about zero during the MPT (Fig. 2b). This suggests the existence of two ‘regimes’: before the MPT, deglaciations of small ice



**Figure 1 | Three-million-year time series of marine oxygen isotope values, subarctic surface air temperature and ice volume.** Thick lines represent 400-kyr running means. **a**, Input marine oxygen isotope ( $\delta^{18}\text{O}$ ) stack record<sup>12</sup> with the present-day value of 3.22‰ subtracted (red), and the reconstructed ice-sheet (blue) and deep-water temperature (green) components. **b**, Reconstructed surface air temperature (annual average

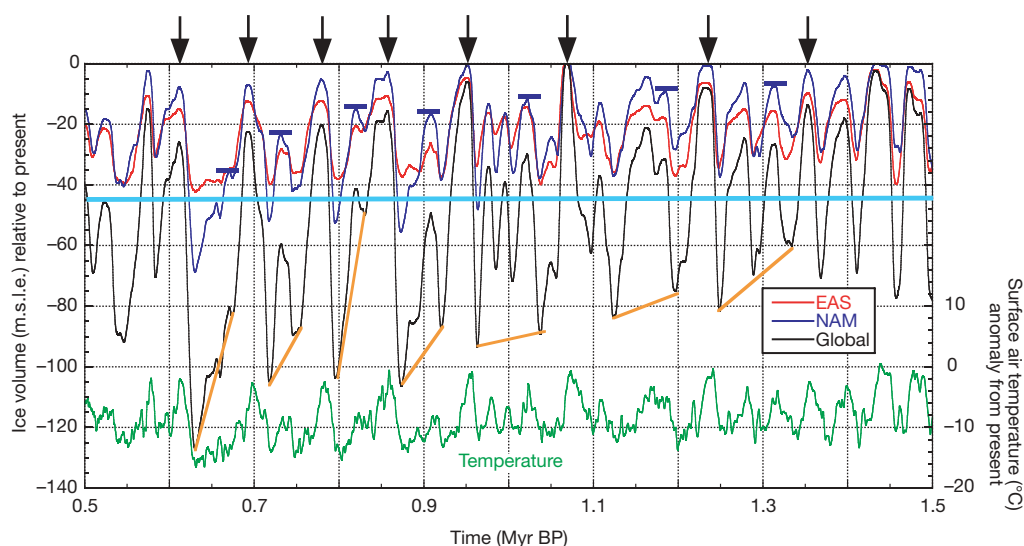
Northern Hemisphere continents between 40° and 80° N) and deep-ocean temperature. **c**, Reconstructed sea level or global ice volume (green), and the Eurasian (red) and North American (blue) contributions. Coloured lines/arrows along top represent various stages of glaciation. MPT, mid-Pleistocene transition; NH, Northern Hemisphere; BP, before present; m.s.l.e., m sea level equivalent.



**Figure 2 | Phase lags of input/reconstructed variables, as evaluated using Blackman–Tukey cross-spectral analysis with 30% lags.** Data points represent 500-kyr windows. **a**, Phase lag at 41 kyr (obliquity) of Northern Hemisphere surface air temperature, Eurasian (EAS) ice volume, North American (NAM) ice volume, global sea level or total ice volume, and total marine oxygen isotope ( $\delta^{18}\text{O}$ ) value, relative to insolation ( $65^\circ\text{N}$ , 21 June). For the obliquity spectral component, the coherence is always above the 95% confidence level for all variables considered, whereas the reconstructed records are generally incoherent with respect to the precession component and also to the eccentricity components (before 1.2 Myr ago). As the tuning of the LR04 stack emphasizes obliquity, the associated lag is primarily determined by the assumed time constant of the simple ice model that was used for tuning<sup>12</sup>. The lag of  $\delta^{18}\text{O}$  relative to insolation increased between 2.5 and 1.5 Myr ago, after which it remained largely constant, as this was prescribed as such in the tuning of the LR04 stack<sup>12</sup>. Lags and leads of all other variables relative to  $\delta^{18}\text{O}$  are independent of this tuning procedure, but are most probably associated with considerable uncertainty owing to the limited availability of palaeoclimate constraints and to the simplicity of mainly the ocean model<sup>11</sup>. **b**, Upper four curves (left axis) show phase lag of global ice volume to Northern Hemisphere surface air temperature for 41-kyr inception and deglaciation stages (light and dark blue) and for 100-kyr inception and deglaciation stages (red and orange). Lower two curves (right axis) show spectral power of 100-kyr Northern Hemisphere surface air temperature (black) and global sea level or ice volume (green) variations.

sheets linearly followed temperature, whereas after the MPT the huge ice sheets were actively involved in the rapid, nonlinear deglaciations to terminate the 100-kyr glacials. Even though they were ultimately triggered by insolation, inceptions were probably governed by non-ice-sheet climate feedbacks to decrease temperature below the threshold of ice-sheet inception, whereas glacial terminations presumably actively involved the rapid break-up of vulnerable large ice sheets<sup>13,24</sup>, both agreeing with phase relationships over the last 360 kyr derived from Antarctic ice cores<sup>25</sup>.

We note that the resulting phasings of reconstructed variables relative to insolation are largely determined by the choices made to construct the age model of LR04 (which determines the phase between  $\delta^{18}\text{O}$  and insolation). However, the 1.5–4.5-kyr lag between reconstructed air temperature and insolation (Fig. 2a) agrees well with the response timescale of the important feedback mechanisms in the climate system (ocean, carbon cycle, ice sheets), suggesting that the phasings assumed in LR04 are realistic. Also, the reconstructed



**Figure 3 | Time series of ice volume and Northern Hemisphere subarctic surface air temperature between 1.5 and 0.5 Myr ago, across the MPT.** Left axis, ice volume; global, Eurasian (EAS) and North American (NAM) contributions are shown as black, red and dark blue curves, respectively. Right axis, Northern Hemisphere subarctic surface air temperature (green curve). Vertical black arrows depict inferred 100-kyr interglacials.

Horizontal dark blue bars represent the first NAM ice volume minimum following a 100-kyr interglacial. The horizontal light blue line represents the 45 m.s.l.e. level beyond which the Cordilleran and Laurentide ice sheets in North America merged (Supplementary Information), which occurred only after 1 Myr ago. Orange lines connect maxima in global ice volume within a single 100-kyr glacial, highlighting increasing ice volumes during a glacial.

lags to insolation vary in time in a manner not prescribed by LR04. These and other considerations (see Supplementary Information) indicate that our analysis provides a picture of the phasings between climate variables and insolation that is consistent overall.

Our results support other analyses in showing that the MPT represents a fairly gradual increase in 100-kyr power (Fig. 2b) from (roughly) 1.4 to 0.6 Myr ago<sup>5,13,26</sup>. An important feature is the increased inability of the system to reach full interglacial levels at the obliquity and precession timescales (Fig. 3), in particular in North America where continuing cooling enabled ice sheets to overcome insolation maxima at ever greater volumes<sup>6,24</sup>. This produced further expansion and eventual merging (at about 45 m.s.l.e.) of the Cordilleran and the Laurentide ice sheets. In Eurasia, the post-MPT ice volume was still limited to about 40 m.s.l.e., presumably because the Atlantic Ocean in the west and the dry and cold Siberian plains to the east prevented further expansion of the Fennoscandian ice sheet<sup>11</sup>.

The appearance of full-grown, merged ice sheets in North America coincided with the deglaciation lag between surface air temperature and ice volume being reduced to about zero (Fig. 2b), suggesting a leading role for the NAM ice sheets in glacial terminations. We argue that this large NAM ice volume initiated rapid deglaciations involving large ice cap instabilities<sup>17,27</sup>, because similar insolation maxima apparently failed to trigger deglaciations of less extensive ice sheets. Thick, wet-based ice sheets are particularly sensitive to climate warming<sup>13,14</sup>. Hence, once the NAM ice sheets reached a certain threshold size (about 70 m.s.l.e.) during the post-MPT glacials, the following insolation maximum and associated air temperature peak triggered a rapid thinning of the ice sheet through increased sliding rates<sup>28</sup>, presumably via fast-flowing ice streams and possibly enhanced by ablation, sea level and ocean circulation feedbacks<sup>6,10,29</sup> (Supplementary Information).

Even though our 3-Myr reconstructions of temperature, ice volume and sea level should be considered first-order estimates—mainly owing to the limited availability of constraining palaeoclimate data—they nevertheless suggest that differences between the NAM and EAS ice sheets were central to the gradual emergence of the 100-kyr glacials between 1.4 and 0.6 Myr ago. Whereas the dry and sluggish EAS ice sheets dominated before the MPT, increased climate cooling enabled the NAM ice sheets to expand and eventually combine into one large ice sheet that ultimately became unstable, triggering rapid glacial terminations. The duration of the post-MPT glacials of 80–120 kyr (2–3 obliquity cycles) was essentially governed by the rate at which the NAM ice sheets expanded. According to this hypothesis, specific ice dynamics and ice–climate interactions related to the North American ice sheets played a crucial role in establishing the 100-kyr glacial cycles.

## METHODS SUMMARY

The key to separating the LR04 stack marine  $\delta^{18}\text{O}$  record<sup>12</sup> into actual climate data is the observation that the two main components affecting the marine  $\delta^{18}\text{O}$  value (storage in ice sheets and deep-water temperature) both rely heavily on surface air temperature (with regard to long-term changes). By using models that describe these relationships (an elaborate ice-sheet model to link air temperature to ice volume and ice  $\delta^{18}\text{O}$  storage, and a simple ocean-temperature model to couple deep-water temperature to surface air temperature) in conjunction with an inverse methodology, we were able to quantify the two components<sup>11</sup> for the 3-Myr LR04 stack marine  $\delta^{18}\text{O}$  record. In the process, 3-Myr-long mutually consistent records of atmospheric temperature, ice volume and global sea level were reconstructed. Using Blackman–Tukey cross-spectral analysis, we estimated the phasings between these climate records in order to assess causes and effects during glacial fluctuation, mainly for inception and termination stages. Even though the Blackman–Tukey method has its drawbacks in terms of accuracy, we decided to use it to compare our phasings with those of LR04 (where it was also used). The main problem with new long records such as those presented here is to find sufficiently long and independent records for validation. A comparison between our reconstructed records and existing independent proxy records is presented in the Supplementary Information.

Received 13 April; accepted 27 May 2008.

1. Raymo, M. E. The initiation of Northern Hemisphere glaciation. *Annu. Rev. Earth Planet. Sci.* **22**, 353–383 (1994).
2. Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. *Science* **287**, 269–272 (2000).
3. Lawrence, K. T., Zhonghui, L. & Herbert, T. D. Evolution of the eastern tropical Pacific through Plio-Pleistocene glaciation. *Science* **312**, 79–83 (2006).
4. Hays, J. D., Imbrie, J. & Shackleton, N. J. Variation in the Earth's orbit: Pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).
5. Clark, P. U. *et al.* The middle Pleistocene transition: Characteristics, mechanisms, and implication for long-term changes in atmospheric  $\text{pCO}_2$ . *Quat. Sci. Rev.* **25**, 3150–3184 (2006).
6. Berger, A., Li, X. S. & Loutre, M. F. Modelling northern hemisphere ice volume over the last 3 Ma. *Quat. Sci. Rev.* **18**, 1–11 (1999).
7. Clark, P. U. & Pollard, D. Origin of the Middle Pleistocene transition by ice sheet erosion of regolith. *Paleoceanography* **13**, 1–9 (1998).
8. Tziperman, E. & Gildor, H. On the mid-Pleistocene transition to 100-kyr glacial cycles and the asymmetry between glaciation and deglaciation times. *Paleoceanography* **18**, 1001, doi:10.1029/2001PA000627 (2003).
9. Muller, R. A. & MacDonald, G. J. Glacial cycles and orbital inclination. *Nature* **377**, 107–108 (1995).
10. Raymo, M. E., Lisiecki, L. E. & Nisancioglu, K. H. Plio-Pleistocene ice volume, Antarctic climate, and the global  $\delta^{18}\text{O}$  record. *Science* **313**, 492–495 (2006).
11. Bintanja, R., van de Wal, R. S. W. & Oerlemans, J. Modelled atmospheric temperatures and global sea level over the past million years. *Nature* **437**, 125–128 (2005).
12. Lisiecki, L. E. & Raymo, M. E. A Pliocene-Pleistocene stack of 57 globally distributed benthic  $\delta^{18}\text{O}$  records. *Paleoceanography* **20**, doi:10.1029/2004PA001071 (2005).
13. Clark, P. U., Alley, R. B. & Pollard, D. Northern Hemisphere ice-sheet influences on global climate change. *Science* **286**, 1104–1111 (1999).
14. Marshall, S. J. & Clark, P. U. Basal temperature evolution of North American ice sheets and implications for the 100-kyr cycle. *Geophys. Res. Lett.* **29**, doi:10.1029/2002GL015192 (2002).
15. Lisiecki, L. E. & Raymo, M. E. Plio-Pleistocene climate revolution: Trend and transitions in glacial cycles. *Quat. Sci. Rev.* **26**, 56–69 (2007).
16. Raymo, M. E. & Nisancioglu, K. H. The 41 kyr world: Milankovitch's other unsolved mystery. *Paleoceanography* **18**, 1011, doi:10.1029/2002PA000791 (2003).
17. Imbrie, J. *et al.* in *Milankovitch and Climate* (ed. Berger, A. L.) 269–305 (Reidel, Dordrecht, 1984).
18. Maslin, M. A., Li, X. S., Loutre, M. F. & Berger, A. The contribution of orbital forcing to the progressive intensification of northern hemisphere glaciation. *Quat. Sci. Res.* **17**, 411–426 (1998).
19. Ruddiman, W. F. Orbital insolation, ice volume and greenhouse gases. *Quat. Sci. Rev.* **22**, 1597–1629 (2003).
20. Shackleton, N. J. The 100,000-year ice-age cycle identified and found to lag temperature, carbon dioxide, and orbital eccentricity. *Science* **289**, 1897–1902 (2000).
21. Maslin, M. A. & Ridgeway, A. J. in *Early-Middle Pleistocene Transitions: The Land-Ocean Evidence* (eds Head, M. J. & Gibbard, P. L.) 19–34 (GSL Special Publication Vol. 247, Geological Society, London, 2006).
22. Valdes, P. & Glover, R. W. Modelling the climate response to orbital forcing. *Phil. Trans. R. Soc. Lond. A* **357**, 1873–1890 (1999).
23. Shackleton, N. J. *et al.* Oxygen isotope calibration of the onset of ice-rafting and history of glaciation in the North Atlantic region. *Nature* **307**, 620–623 (1984).
24. Raymo, M. E. The timing of major climate terminations. *Paleoceanography* **12**, 577–585 (1997).
25. Kawamura, K. *et al.* Northern Hemisphere forcing of climatic cycles in Antarctica over the past 360,000 years. *Nature* **448**, 912–916 (2007).
26. Huybers, P. Glacial variability over the last two million years: An extended depth-derived age model, continuous obliquity pacing, and the Pleistocene progression. *Quat. Sci. Rev.* **26**, 37–55 (2007).
27. Weertman, J. Stability of the junction between an ice sheet and an ice shelf. *J. Glaciol.* **13**, 3–11 (1974).
28. Parrenin, F. & Paillard, D. Amplitude and phase of glacial cycles from a conceptual model. *Earth Planet. Sci. Lett.* **214**, 243–250 (2003).
29. Alley, R. B. & Clark, P. U. The deglaciation of the Northern Hemisphere: A global perspective. *Annu. Rev. Earth Planet. Sci.* **27**, 149–182 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** Financial support to R.B. was provided by the Netherlands Organisation of Scientific Research (NWO), in the framework of the SPINOZA award of J. Oerlemans, and through the EU programme Quantify. This work was initiated during the start-up phase of the Utrecht Centre for Geosciences (UCG) programme. We thank W. Greuell, D. Lea, L. Lourens, M. Siddall and N. Weber for remarks on earlier versions of the paper.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to R.B. (bintanja@knmi.nl).

# Inner-core shear-wave anisotropy and texture from an observation of PKJKP waves

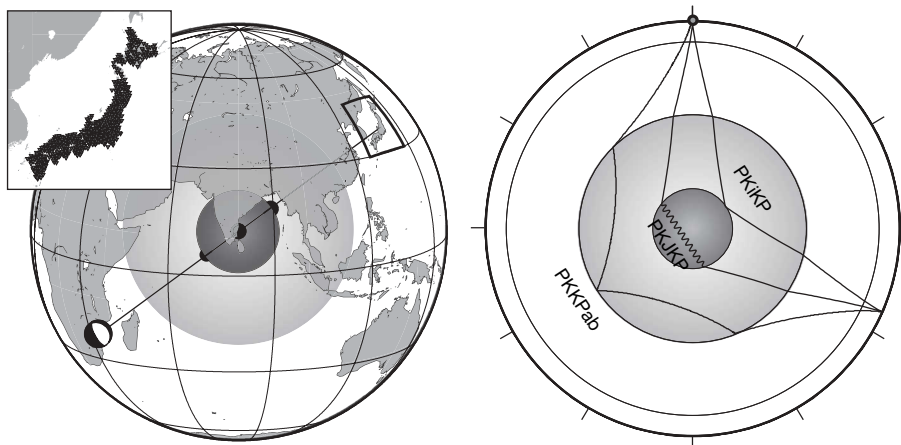
James Wookey<sup>1</sup> & George Helffrich<sup>1</sup>

Since the discovery of the Earth's core a century ago<sup>1</sup>, and the subsequent discovery<sup>2</sup> of a solid inner core (postulated to have formed by the freezing of iron<sup>3</sup>) seismologists have striven to understand this most remote part of the deep Earth. The most direct evidence for a solid inner core would be the observation of shear-mode body waves that traverse it, but these phases are extremely difficult to observe. Two reported observations in short-period data<sup>4,5</sup> have proved controversial<sup>6</sup>, but such averaging limits the usefulness of the observations to reported sightings. We present two observations of an inner-core shear-wave phase at higher frequencies in stacked data from the Japanese High-Sensitivity Array, Hi-Net<sup>8</sup>. From an analysis of timing, amplitude and waveform of the 'PKJKP' phase we derive constraints on inner-core compressional-wave velocity and shear attenuation at about 0.3 Hz which differ from standard isotropic core models<sup>9</sup>. We can explain waveform features and can partially reconcile the otherwise large differences between core wavespeed and attenuation models that our observations apparently suggest if we invoke shear-wave anisotropy in the inner core. A simple model of an inner core composed of hexagonal close-packed iron with its *c* axis aligned perpendicular to the rotation axis<sup>10</sup> yields anisotropy that is compatible with both the shear-wave anisotropy that we observe and the well-established 3 per cent compressional-wave anisotropy.

Our primary observation comes from a single event which occurred on 22 February 2006 in Mozambique, at an epicentral distance of 113.7° degrees from the centre of Hi-Net<sup>8</sup> (Fig. 1). This was a shallow event with a moment magnitude  $M_w = 7.0$ . For such a large,

shallow event the source time function is remarkably impulsive (see Supplementary Fig. 1). Phase-weighted stacks<sup>11</sup> (Fig. 2) show very clear PKKPab, pPKKPab, PKiKP and pPKiKP arrivals, and weak PKKPbc and pPKKPbc. The separation between the primary arrivals and the depth phases best fit an event 14 km deep. Figure 2e shows the time–slowness window for PKJKP as predicted by the Earth model ak135 (ref. 9) and a clear arrival is observed. We are also able to match the arrival with synthetics (Fig. 3). As this phase arrives very close to the ak135 prediction, we suggest that this Earth model provides an accurate average inner-core shear-wave velocity. We also have made a second observation of PKJKP, from a 2007 event on the mid-Atlantic ridge, but a complex wavetrain precludes further useful analysis (see Supplementary Figs 7–8).

The PKJKP waveform presented in Fig. 3 shows another interesting feature: an arrival about 7 s after PKJKP (and 3 s after pPKJKP). This cannot be sPKJKP or another near-source converted phase, as no such phase is observed for PKKPab or PKiKP, which have very similar upper-mantle ray paths (Fig. 2). Furthermore, sPKJKP would leave a shallow event almost vertically, so the conversion to a P-wave at the free surface would be extremely weak. We interpret this secondary phase as the effect of inner-core seismic anisotropy (see Supplementary Information). Seismic anisotropy is the variation of seismic wavespeed with direction, which has been observed for the inner core using P-wave and normal mode data (see, for example, ref. 12 and references therein). When a shear-wave encounters an anisotropic medium it is split into two components whose polarizations are defined by the symmetry of the medium, and which will separate in time as they propagate through the region. In our scheme, this would lead to two distinct PKJKP phases (which we denote PKJKP1



**Figure 1 | Source, ray path and receiver geometry.** We searched for evidence of PKJKP in records at the Japanese Hi-Net array (inset) of the  $M_w = 7.0$  shallow event (depth ~14 km) on 22 February 2006 in

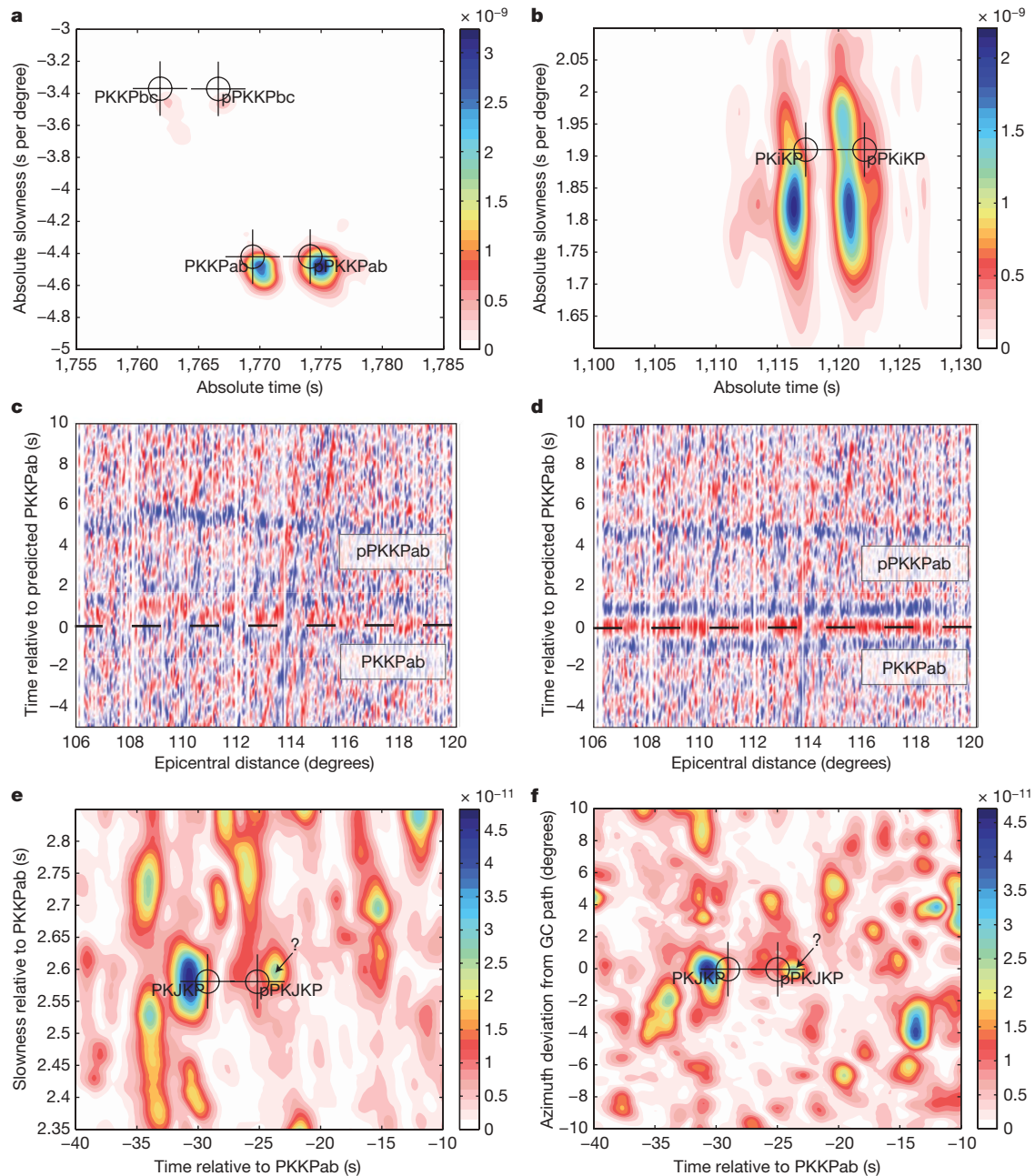
Mozambique. The epicentral distance to the centre of the array is 113.7°. The right panel shows the ray paths for PKKPab, PKiKP and PKJKP at this distance (straight lines are P-wave segments, wavy lines are S-wave).

<sup>1</sup>Department of Earth Sciences, University of Bristol, Wills Memorial Building, Queens Road, BS8 1RJ, UK.

and PKJKP2). This has been predicted<sup>12,13</sup>, but never before observed. Figure 3 shows the degree of splitting expected for shear-wave anisotropy between 0 and 2%; to explain the delay we observe requires an anisotropy of  $\sim 1\%$  averaged across the propagation path through the inner core.

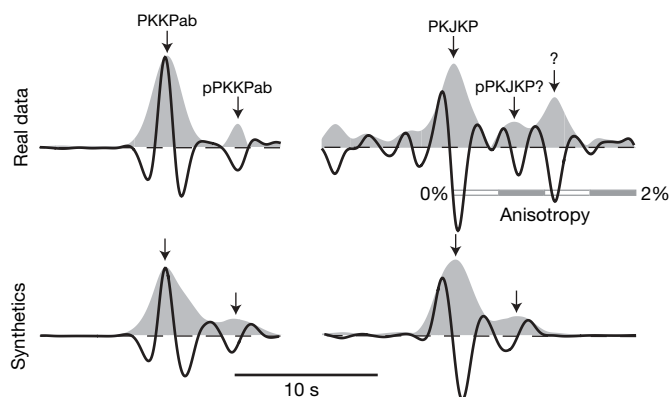
The simplest model proposed to explain the bulk anisotropy of the inner core is the alignment of iron crystals<sup>12,13</sup>. It has even been suggested that the inner core might be a single crystal of hexagonal close-packed (h.c.p.) iron<sup>13</sup>. The anisotropy that we measure is small compared with those predicted for single-crystal iron at inner-core

conditions<sup>10,14</sup> (up to  $\sim 20\%$ ; see Supplementary Fig. 10). To explain our inferred shear-wave anisotropy as well as the much better-established P-wave anisotropy, we have explored a range of simple textural models for these elastic constants. The models are a series of orientations and rotational averages of h.c.p. and body-centred cubic (b.c.c.) iron, to provide an estimate of the aggregate elasticity (and hence anisotropy). These models are motivated by different ideas of crystal alignment due to dendritic growth during solidification<sup>15,16</sup> or post-solidification deformation<sup>15,17,18</sup>. Figure 4 outlines these mechanisms. Models tested are given in Supplementary Table 1



**Figure 2 | Seismic data.** **a, b**, Vespagrams for the PKKPab and PKiKP time–slowness windows, respectively. The colour scale represents the amount of energy across all the traces at a given time and slowness. These vespagrams are calculated using a phase-weighted slant stack<sup>11</sup>. Crosshairs denote predicted times and slowness from ak135 (ref. 9) for various core phases. Clear maxima associated with PKKPab, pPKKPab, PKiKP and pPKiKP arrivals are visible, with weaker maxima for PKKPbc and pPKKPbc. **c**, Time window for PKKPab in the unstacked data. Because PKKPab is clearly visible we use it as reference phase to calculate a receiver-side static time correction<sup>30</sup>, and **d** shows this correction applied to PKKPab.

**e**, Time–slowness window (relative to the PKKPab reference phase) where PKJKP is predicted to arrive. A clear maximum can be seen  $\sim 1.5$  s before the prediction, at the correct slowness to within the resolution of the array ( $\sim 0.05$  s per degree). There is also energy near the time predicted for pPKJKP, although this is low amplitude (near the noise level) and poorly constrained in slowness. **f**, Azimuthal slant stack, at a fixed slowness of 2.6 s per degree. This shows that the maximum identified as PKJKP arrives within  $0.2^\circ$  of the (major arc) great circle path. A second peak is also seen at the slowness of PKJKP (denoted by the question mark).

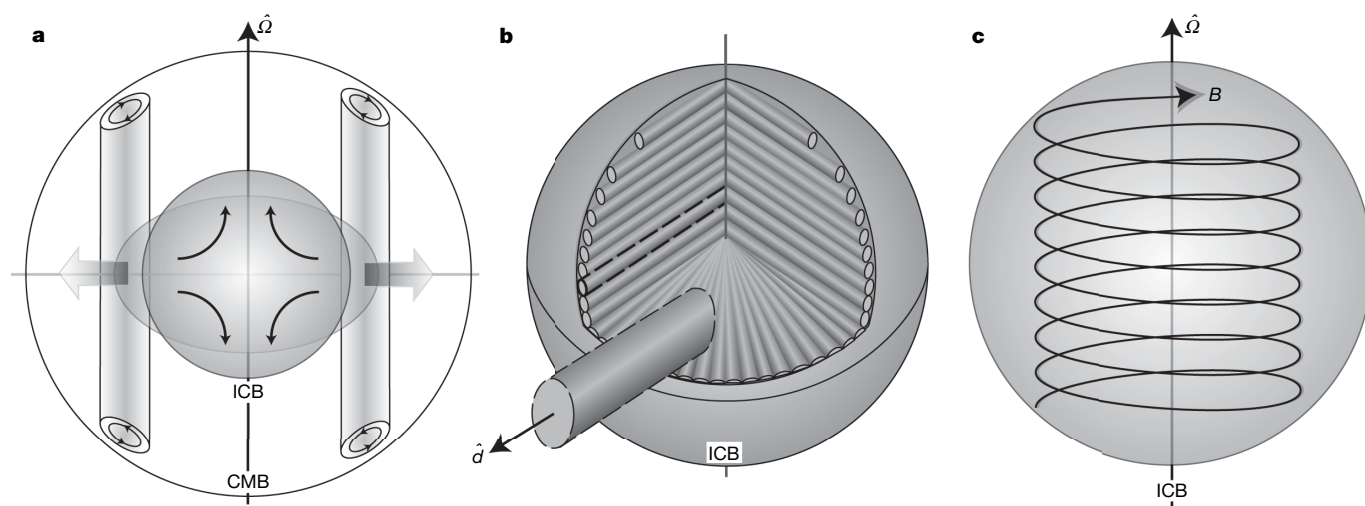


**Figure 3 | Real (top) and synthetic waveforms (bottom).** These are phase-weighted slant stacks at the peak slownesses for PKKPab and PKJKP. The shaded area is the envelope function of the trace. Given the uncertainty in relative amplitudes from moment tensor solutions<sup>32</sup>, in order to match the observed amplitude ratio of between PKKPab and pPKKPab in the real data, a taper of value 0.25 is applied to the PKKPab synthetics after the primary arrival. The same taper is also applied to the PKJKP synthetics. The PKJKP phase observed is the Hilbert transform of the reference PKKPab phase, as predicted. An arrival is also apparent near the correct time for pPKJKP, although this is very close to the noise level of the trace and not well constrained. The arrival denoted by the question mark may be due to shear-wave splitting by inner-core anisotropy. The scale bar shows the time lag predicted for a model of uniform anisotropy, implying that  $\sim 1\%$  is required to explain this phase.

and Supplementary Fig. 11. A first-order picture of inner-core anisotropy shows that P-wave phases on polar raypaths traverse the region  $\sim 3\%$  faster than those on equatorial paths<sup>12</sup>. The only models tested that even roughly match this constraint involve the rotational average of h.c.p. Fe with the  $c$  axis perpendicular to the Earth's rotation axis<sup>10</sup>. These models also show a shear-wave anisotropy that is roughly compatible with that inferred from our observation of PKJKP. Thus, the measured anisotropy is currently non-unique with regard to its cause (that is, depositional or deformational) but does suggest iron with h.c.p. structure to be a better prospect than iron with b.c.c. structure. Previous observations of inner-core shear-wave anisotropy come from inversions of normal-mode observations<sup>19,20</sup>, which

predict larger absolute anisotropies. The 7-s splitting we observe for PKJKP is, in fact, compatible with these (see Supplementary Information), as such models predict a change in sign of anisotropy at a radius between 400 and 800 km (see ref. 20), leading to a relatively small aggregate anisotropy along the ray path. Mechanisms to explain such variation in anisotropy, however, are more complex than the simple textural models proposed above.

Also of interest is the strength of the PKJKP arrival, which we measure using the ratio of the recorded amplitudes PKJKP and PKiKP (hereafter  $R_{j/i}$ ) in linear slant stacks. In our data,  $R_{j/i} = 0.14 \pm 0.05$ , significantly larger than  $R_{j/i} = 0.02$  predicted using reflectivity modelling for a model compatible with ak135. This could be due to either an increased PKJKP amplitude or a reduced PKiKP amplitude. Inner-core anisotropy might affect PKJKP and PKiKP amplitudes by altering the reflection and coefficients at the inner-core boundary (ICB). For example we find, using elastic constants from ref. 14, that the ratio of PKJKP to PKiKP amplitude can vary significantly: up to a factor of 25 depending on the degree of alignment and orientation at the ICB (see Supplementary Fig. 14). This has also been suggested to explain observations of anomalous amplitudes of PKPdf phases<sup>21</sup>. As this effect alone explains both waveform features and the observed  $R_{j/i}$  without requiring modification of the core's isotropic properties, we favour it. But to test whether we can in fact reconcile our observation with an isotropic, spherically symmetric model of the inner core, we have also run a suite of synthetics varying the seismic velocity ( $V_p$  and  $V_s$ ), density and  $Q_s$  (the shear-wave 'quality' factor, also sometimes called  $Q_\mu$ ) in the inner core (see Supplementary Information). We find that  $R_{j/i}$  depends most strongly on  $V_p$  of the inner core at the ICB (predominantly through the amplitude of PKiKP), with a weaker but significant dependence on  $Q_s$  (through the amplitude of PKJKP). The influence of  $V_s$  and density (at least within bounds compatible with normal-mode data) is negligibly weak. Supplementary Fig. 13 shows the variation in  $R_{j/i}$  with inner-core  $V_p$  and  $Q_s$ , identifying the distribution of models consistent with our measured  $R_{j/i}$ . This shows that to reconcile our measured  $R_{j/i}$  with an isotropic one-dimensional model of the inner core requires a much higher  $Q_s$  than is inferred from normal-mode studies, a considerably reduced  $V_p$  (at least near the ICB), or a more moderate combination of the two. For example, in order that  $Q_s$  be compatible with the normal-mode



**Figure 4 | Possible causes of crystal alignment in the inner core.** **a**, Compressional and tensional deformation due to non-uniform growth. Outer-core convection in Taylor columns<sup>33</sup> leads to larger equatorial heat-flux, promoting freezing at the ICB in these regions. This oblateness is dynamically unstable, leading to deformation<sup>17</sup> symmetrically about the rotation axis ( $\Omega$ ). CMB, core-mantle boundary. **b**, Crystal alignment due to dendritic solidification. As liquid iron freezes onto the ICB, dendrite

structures might be formed<sup>16</sup> and could persist deep into the inner core. In this case crystals would be oriented relative to the dendrite long axes ( $d$ ), which we assume to be perpendicular to  $\Omega$ . Solid-state, inward axial flow modifies orientation (Supplementary Table 1 and Fig. 11). **c**, Alignment due to Maxwell stresses (after ref. 18). In this model, stresses exerted by the Earth's magnetic field ( $B$ ) re-orient crystals of inner-core iron, leading to large-scale texturing.

measurements of ref. 22, the P-wave velocity increase at the ICB must be of the order of  $0.3 \text{ km s}^{-1}$  (compared with  $0.75 \text{ km s}^{-1}$  in ak135). A global degree-1 asymmetry of P-wave speed (with a deviation from the reference model of  $-0.7$  to  $+0.2\%$ ) in the outermost part of the inner core has been documented<sup>23,24</sup>, but this variation is much smaller than our observations require. A strong variation in PKiKP amplitude has also been observed by using P or PcP as a reference phase<sup>25–27</sup>. It is also possible, however, that attenuation values derived from normal modes are not appropriate at body-wave frequencies (see refs 28, 29 and Supplementary Information). The amplitude spectrum of the PKJKP (see Supplementary Fig. 9) also suggests that  $Q_s$  in the inner core might be higher than previously estimated. This is hard to reconcile with measurements of P-wave  $Q$  from PKIKP<sup>28,29</sup>, without significant bulk attenuation, and perhaps a layered  $Q$  structure in the inner core. The influence of scattering attenuation may also be important<sup>29</sup>. These issues are discussed further in the Supplementary Information, but drawing firm conclusions as to the causes will require more observations than provided by a single measurement. There are several factors that might influence the amplitudes of PKJKP and PKiKP beyond those that may be tested by our modelling methodology, including the focusing/defocusing effects of ICB topography or inner-core velocity heterogeneity. However, we believe that the simplest explanation remains the effect of anisotropy.

To test these ideas further requires more PKJKP observations at different traversal directions across the inner core. We believe that as the new generation of large-aperture dense array experiments such as Hi-Net and the nascent USArray begin to accumulate data, PKJKP should become a more routinely detected phase. Such accumulated observations will probe other areas of the inner core to assess the possible hemispheric variation in its shear anisotropy; our observations are confined to the quasi-western hemisphere<sup>24</sup>. This will provide stronger constraints on the depth extent of texturing and inner-core growth through time.

## METHODS SUMMARY

Our data comprise 704 short-period Hi-Net<sup>8</sup> records from borehole seismometers sited around the Japanese Islands. We use PKKPab as a reference phase to calculate a receiver-side static time correction<sup>30</sup>. Data are initially bandpass filtered between 0.05 and 2 Hz, to remove high-frequency noise. A power-2 phase-weighted slant stack<sup>11</sup> is applied and the data are bandpass filtered again between 0.05 and 0.5 Hz to smooth the spectra, and a trace envelope function is derived. Figure 2e shows the time–slowness window for PKJKP as predicted by ak135 (ref. 9). A clear energy pulse can be observed at a slowness which is indistinguishable from the model-predicted slowness, and  $\sim 1.5 \text{ s}$  earlier than the predicted arrival time. Azimuthal stacking shows a deviation of  $<0.2^\circ$  from the major arc great circle path (Fig. 2f). We interpret this phase to be PKJKP. We are also able to match the arrival with reflectivity method<sup>31</sup> synthetics for a one-dimensional isotropic Earth model<sup>9</sup> (Fig. 3). The PKJKP phase observed is the Hilbert transform of the reference PKKPab phase, as predicted. Further tests are detailed in the Supplementary Information. For amplitude measurements, linear slant stacks are used to avoid biasing by phase weighting. Amplitude error bounds are estimated from the signal-to-noise ratio of the PKJKP peak; those of PKiKP are negligible in comparison. A correction is also made for the moment tensor.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 22 November 2007; accepted 28 May 2008.**

1. Oldham, R. D. The constitution of the interior of the Earth, as revealed by earthquakes. *Q. J. Geol. Soc. Lond.* **62**, 456–475 (1906).
2. Lehmann, I. P'. *Bur. Centr. Sismol. Int.* **A 14**, 87–115 (1936).
3. Birch, F. The alpha–gamma transformation of iron at high pressures, and the problem of the Earth's magnetism. *Am. J. Sci.* **238**, 192–211 (1940).
4. Julian, B. R., Sheppard, R. M. & Davies, D. PKJKP. *Nature* **235**, 317–318 (1972).
5. Okal, E. A. & Cansi, Y. Detection of PKJKP at intermediate periods by progressive multi-channel correlation. *Earth Planet. Sci. Lett.* **164**, 23–30 (1998).
6. Deuss, A., Woodhouse, J. H., Paulssen, H. & Trampert, J. The observation of inner core shear waves. *Geophys. J. Int.* **142**, 67–73 (2000).

7. Cao, A. M., Romanowicz, B. & Takeuchi, N. An observation of PKJKP: Inferences on inner core shear properties. *Science* **308**, 1453–1455 (2005).
8. Obara, K., Kasahara, K., Hori, S. & Okada, Y. A densely distributed high-sensitivity seismograph network in Japan: Hi-net by National Research Institute for Earth Science and Disaster Prevention. *Rev. Sci. Instrum.* **76**, 21301–1–12 (2005).
9. Kennett, B. L. N., Engdahl, E. R. & Buland, R. Constraints on seismic velocities in the Earth from traveltimes. *Geophys. J. Int.* **122**, 108–124 (1995).
10. Steinle-Neumann, G., Stixrude, L., Cohen, R. E. & Gülseren, O. Elasticity of iron at the temperature of the Earth's inner core. *Nature* **413**, 57–60 (2001).
11. Schimmel, M. & Paulssen, H. Noise reduction and detection of weak, coherent signals through phase-weighted stacks. *Geophys. J. Int.* **130**, 497–505 (1997).
12. Song, X. Anisotropy of the Earth's inner core. *Rev. Geophys.* **35**, 297–314 (1997).
13. Stixrude, L. & Cohen, R. E. High-pressure elasticity of iron and anisotropy of Earth's inner core. *Science* **267**, 1972–1975 (1995).
14. Vočadlo, L. Ab initio calculations of the elasticity of iron and iron alloys at inner core conditions: evidence for a partially molten inner core? *Earth Planet. Sci. Lett.* **254**, 227–232 (2007).
15. Barrett, C. & Massalski, T. *Structure of Metals: Crystallographic Methods Principles and Data* 3rd rev. edn (Pergamon, Oxford, 1980).
16. Bergman, M. I., MacLeod-Silberstein, M., Haskel, M., Chandler, B. & Akpan, N. A laboratory model for solidification of Earth's core. *Phys. Earth Planet. Int.* **153**, 150–164 (2005).
17. Yoshida, S., Sumita, I. & Kumazawa, M. Growth model of the inner core coupled with the outer core dynamics and the resulting elastic anisotropy. *J. Geophys. Res.* **101**, 28085–28103 (1996).
18. Buffett, B. A. & Wenk, H. R. Texturing of the Earth's inner core by Maxwell stresses. *Nature* **413**, 60–63 (2001).
19. Tromp, J. Support for anisotropy of the Earth's inner core from free oscillations. *Nature* **366**, 678–681 (1993).
20. Beghein, C. & Trampert, J. Robust normal mode constraints on inner-core anisotropy from model space search. *Science* **299**, 552–555 (2003).
21. Creager, K. Anisotropy of the inner core from differential travel-times of the phases PKP and PKiKP. *Nature* **356**, 309–314 (1992).
22. Widmer, R., Masters, G. & Gilbert, F. Spherically symmetric attenuation within the Earth from normal mode data. *Geophys. J. Int.* **104**, 541–553 (1991).
23. Kaneshima, S. Mapping heterogeneity of the uppermost inner core using two pairs of core phases. *Geophys. Res. Lett.* **23**, 3075–3078 (1996).
24. Tanaka, S. & Hamaguchi, H. Degree one heterogeneity and hemispherical variation of anisotropy in the inner core from PKP(BC)–PKP(DF) times. *J. Geophys. Res.* **102**, 2925–2938 (1997).
25. Koper, K. D. & Pyle, M. L. Observations of PKiKP/PcP amplitude ratios and implications for Earth structure at the boundaries of the liquid core. *J. Geophys. Res.* **109**, B03301 (2004).
26. Krasnoshekov, D. N., Kaazik, P. B. & Ovtchinnikov, V. M. Seismological evidence for mosaic structure of the surface of the Earth's inner core. *Nature* **435**, 483–487 (2005).
27. Koper, K. D. & Dombrovskaya, M. Seismic properties of the inner core boundary from PKiKP/P amplitude ratios. *Earth Planet. Sci. Lett.* **237**, 680–694 (2005).
28. Romanowicz, B. & Durek, J. J. in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale* (eds Karato, S., Forte, A., Liebermann, R., Masters, G. & Stixrude, L.) *Geophys. Monogr.* **117**, 161–179 (Am. Geophys. Un., Washington DC, 2000).
29. Cormier, V. F. & Li, X. Frequency-dependent seismic attenuation in the inner core 2. A scattering and fabric interpretation. *J. Geophys. Res.* **107**, 2362 (2002).
30. Rawlinson, N. & Kennett, B. L. N. Rapid estimation of relative and absolute delay times across a network by adaptive stacking. *Geophys. J. Int.* **157**, 332–340 (2004).
31. Mueller, G. The reflectivity method; a tutorial. *J. Geophys.* **58**, 153–174 (1985).
32. Helffrich, G. R. How good are routinely determined focal mechanisms? Empirical statistics based on a comparison of Harvard, USGS and ERI moment tensors. *Geophys. J. Int.* **131**, 741–750 (1997).
33. Zhang, K. Nonlinear magnetohydrodynamic convective flows in the Earth's fluid core. *Phys. Earth Planet. Inter.* **111**, 93–103 (1999).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank L. Vočadlo, M. Kendall, M. Bergman and A. Deuss for discussions, C. Beghein for providing her inner-core anisotropy models and L. Stixrude for a review. J.W. was supported by a NERC postdoctoral fellowship grant.

**Author Contributions** Data analysis and modelling was undertaken by J.W., initiated by, and in consultation with, G.H. Phase-weighted stacking code and inner-core texture models were developed by G.H. Manuscript and figures were produced by J.W.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.W. (j.wooke@bristol.ac.uk).

## METHODS

Our data comprise 704 short-period Hi-Net<sup>8</sup> records from borehole seismometers sited around the Japanese Islands. As the PKKPab arrival is very clear, we use it as a reference phase to calculate a receiver-side static time correction<sup>30</sup>. Data are initially bandpass filtered between 0.05 and 2 Hz, to remove high-frequency noise. A power-2 phase-weighted slant stack<sup>11</sup> is applied and the data are bandpass filtered again between 0.05 and 0.5 Hz to smooth the spectra. We assume a linear moveout between PKJKP and PKKPab in slant stacking. The maximum deviation from this is of the order of 0.1 s, much less than the dominant period of the phase, and may be neglected. Finally the trace envelope function is derived. Figure 2e shows the time–slowness window for PKJKP as predicted by ak135 (ref. 9). A clear energy pulse can be observed at a slowness which is indistinguishable from the model-predicted slowness (we calculate the slowness resolution of the array to be about 0.05 s per degree; see Supplementary Fig. 2), and  $\sim 1.5$  s earlier than the predicted arrival time. We interpret this phase to be PKJKP. Other phases that have previously been identified as problematic (for example, PcPPKiKP, PcPPKiKP and PKKPdf; see refs 4, 5, 7) are predicted to arrive well separated in time or slowness or both. Further evidence for the arrival being PKJKP comes from slant stacking with varying azimuth, with a fixed slowness of 2.6 s per degree relative to PKKPab (Fig. 2f). This shows that the arrival has maximum amplitude at a deviation of less than  $0.2^\circ$  from the great circle path and that it arrives on the major arc. We are also able to match the arrival with synthetics. We use reflectivity method<sup>31</sup> synthetics for a one-dimensional isotropic Earth model<sup>9</sup>. These are processed identically to real data, except a static correction is unnecessary. Figure 3 shows the comparison of the slant-stacked waveform with synthetics. The PKJKP phase observed is the Hilbert transform of the reference PKKPab phase, as is predicted by the synthetics. A weak phase is also observed at the correct time for a source 14 km deep as predicted by PKKPab and PKiKP. Further tests are detailed in the Supplementary Information. To compare the shear-wave splitting we infer with previous, normal-mode derived models we calculate<sup>34</sup> accrued lag-time between SH and SV phases, assuming no perturbation of the isotropic ray path. This is done for the minimum, maximum and mean models of ref. 20. We find that the 7-s splitting we measure is compatible with the range of normal-mode models (see Supplementary Information). For amplitude measurements, linear slant stacks are used to avoid biasing by phase weighting. Amplitude error bounds are estimated from the signal-to-noise ratio of the PKJKP peak; those of PKiKP are an order of magnitude smaller because of a much larger amplitude. A correction is also made for the moment tensor.

34. Thomsen, L. Weak elastic anisotropy. *Geophysics* **51**, 1954–1966 (1986).

# Inapparent infections and cholera dynamics

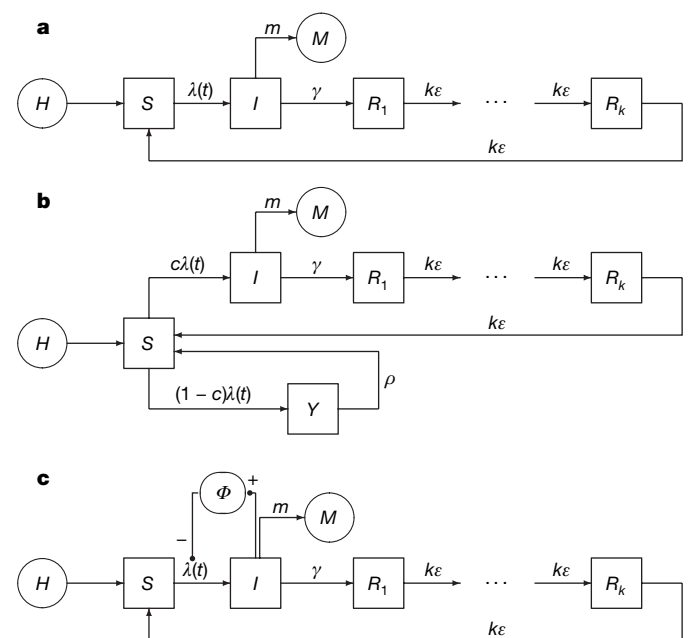
Aaron A. King<sup>1,2</sup>, Edward L. Ionides<sup>3</sup>, Mercedes Pascual<sup>1,4</sup> & Menno J. Bouma<sup>5</sup>

In many infectious diseases, an unknown fraction of infections produce symptoms mild enough to go unrecorded, a fact that can seriously compromise the interpretation of epidemiological records. This is true for cholera, a pandemic bacterial disease, where estimates of the ratio of asymptomatic to symptomatic infections have ranged from 3 to 100 (refs 1–5). In the absence of direct evidence, understanding of fundamental aspects of cholera transmission, immunology and control has been based on assumptions about this ratio and about the immunological consequences of inapparent infections. Here we show that a model incorporating high asymptomatic ratio and rapidly waning immunity, with infection both from human and environmental sources, explains 50 yr of mortality data from 26 districts of Bengal, the pathogen's endemic home. We find that the asymptomatic ratio in cholera is far higher than had been previously supposed and that the immunity derived from mild infections wanes much more rapidly than earlier analyses have indicated. We find, too, that the environmental reservoir<sup>5,6</sup> (free-living pathogen) is directly responsible for relatively few infections but that it may be critical to the disease's endemicity. Our results demonstrate that inapparent infections can hold the key to interpreting the patterns of disease outbreaks. New statistical methods<sup>7</sup>, which allow rigorous maximum likelihood inference based on dynamical models incorporating multiple sources and outcomes of infection, seasonality, process noise, hidden variables and measurement error, make it possible to test more precise hypotheses and obtain unexpected results. Our experience suggests that the confrontation of time-series data with mechanistic models is likely to revise our understanding of the ecology of many infectious diseases.

Cholera is a diarrhoeal disease caused by enteric infection with the bacterium *Vibrio cholerae*. Six of the seven cholera pandemics that have swept the globe since 1817 originated in the low-lying, densely populated regions north of the Bay of Bengal, where the disease is endemic. Although much attention has been focused on cholera<sup>1,8</sup>, unsolved puzzles remain about its mode of transmission and the role of host immunity in its dynamics. This is largely because, in regions where cholera is endemic, most cholera cases are mild or asymptomatic but the true extent of asymptomatic infection has been difficult to assess. Estimates of the ratio of asymptomatic to symptomatic cases vary greatly, and the importance of inapparent infections in the dynamics of cholera outbreaks is unknown. To determine what role is played by inapparent infections, we used an approach that allows indirect inference about unobserved variables.

A remarkably rich data set on the pattern of cholera epidemics exists in the form of mortality records kept by the sanitary commissioners of the former British East Indian province of Bengal<sup>9</sup>. The data consist of monthly cholera death counts in each of 26 districts over the period 1891–1940 (Supplementary Fig. 1). To analyse these data, we formulated a series of models incorporating known or hypothesized mechanisms of transmission and immunity. A

parsimonious model for cholera dynamics is of susceptible–infectious–recovered–susceptible (SIRS) form (Fig. 1a). A novel feature of this model is that it incorporates both transmission tied to human prevalence (using a traditional mass-action term) and transmission from an environmental reservoir (where the pathogen is commonly living in aquatic environments)<sup>5,6,10–12</sup>. This model is a



**Figure 1 | The mechanistic models used.** **a**, SIRS model; **b**, two-path model; **c**, environmental-phage model. Births, related to the total population size  $H$ , are assumed to feed the pool of susceptibles,  $S$ . Individuals are susceptible to infection when born. Exposure to the pathogen occurs at time-dependent rate  $\lambda(t)$ .  $c$  is the probability that an exposure leads to a contagious infection (class  $I$ ). Note that when  $c = 1$  and  $\rho = \infty$ , the two-path model (**b**) reduces to the SIRS model (**a**); when  $c < 1$ , some exposures result in short-term immunity (class  $Y$ ). Infected individuals die at an excess rate  $m$  and recover at a rate  $\gamma$ ; the time an individual spends within the  $I$  class is exponentially distributed. We assume that an individual remains immune to reinfection for a duration gamma-distributed with mean  $1/\epsilon$  and variance  $1/k\epsilon^2$ . Once immunity has waned, an individual re-enters the susceptible pool ( $S$ ). The measured variable is monthly deaths,  $M$ . The mean duration of short-term immunity is  $1/\rho$ . Individuals in each class are subject to constant background mortality at rate  $0.02 \text{ yr}^{-1}$ . The force of infection,  $\lambda(t)$ , includes terms for environmental and human sources of infection and is assumed to vary seasonally. Because the seasonality of cholera dynamics in Bengal is complex, we used a semi-mechanistic approach: transmission was modelled by a flexible periodic function of time. In the environmental-phage model (**c**), as infected hosts shed pathogen, phage  $\Phi$  builds up in the environment and reduces transmissibility. The equations specifying these models are given in the Supplementary Equations.

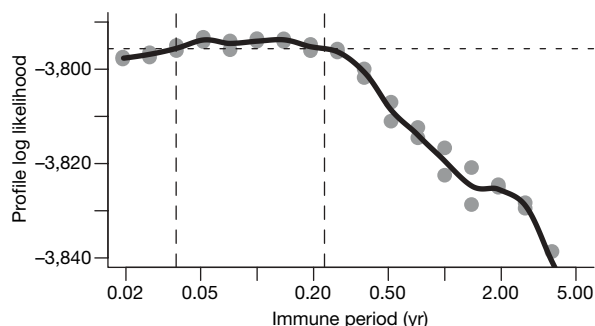
<sup>1</sup>Department of Ecology and Evolutionary Biology, <sup>2</sup>Department of Mathematics, <sup>3</sup>Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>4</sup>Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA. <sup>5</sup>Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, University of London, London WC1E 7HT, UK.

partially observed, nonlinear, stochastic dynamical system in continuous time and as such is not amenable to analysis by standard statistical techniques. Inference for such systems has recently been facilitated by a new likelihood maximization procedure<sup>7</sup> which allows for measurement error, non-stationarity, irregular sampling intervals and the mechanistic inclusion of covariates (see Methods). We fitted the SIRS model to each district's data, assessed its explanatory power using log likelihood and the Akaike information criterion (AIC<sub>c</sub>) and compared it with a semi-mechanistic time-series model for cholera<sup>13</sup> and two non-mechanistic time-series models.

The SIRS model explained the data dramatically better than the previous best fit<sup>13</sup> ( $\Delta\text{AIC}_c > 58$ ; Supplementary Table 1). Parameter estimates reveal some surprises. A prediction of the model, robustly consistent across districts, is that immunity must wane on a timescale of weeks to months (Fig. 2 and Supplementary Table 3). This is in stark contrast to the widely held belief that infection-derived immunity to cholera wanes on a timescale of 3–10 yr (refs 13–17). The model also robustly predicts low case fatality ( $0.004 \pm 0.002$ , mean  $\pm$  s.d. across 26 districts). Because fatality among hospitalized cases was historically in excess of 50%<sup>18</sup>, this implies a very large asymptomatic ratio. The implication is that most exposures result not in severe cholera, but in mild or asymptomatic infection, the immunological memory of which is short-lived. This prediction of a high asymptomatic ratio is consistent with the only intensive field studies of inapparent infections of which we are aware<sup>3,4</sup>. However, because in this model all infected individuals are equally infectious, the prediction is that the vast majority of infectives are 'silent shedders'—infectious but without symptoms. The evidence for such a conclusion is mixed but weak at best<sup>2</sup>. Moreover, because in this model all infections are alike, it is unclear whether the improved dynamical description afforded by the SIRS model over the previous best fit is due to the higher proportion of asymptomatic infections or to the shorter duration of immunity.

To tease apart the immunological consequences of exposure to the pathogen from the degree of infectiousness induced, and to allow for heterogeneity in infectiousness, we formulated a second mechanistic model in which exposed individuals can follow an alternate pathway, deriving immunity to infection while shedding a negligible quantity of vibrio and suffering a negligible disease-induced mortality (Fig. 1b). In essence, this model allows for the possibility that exposure to the pathogen effectively vaccinates against reinfection. We leave it to the data to determine how frequently such natural vaccination occurs and how long the resulting immunity lasts.

Under the two-path model, the timescale of short-term immunity ( $9.9 \pm 4.7$  weeks, mean  $\pm$  s.d. across 26 districts) agrees with that of the SIRS model. The estimated case fatality under this model is  $0.34 \pm 0.21$ —consistent with conditions during this period<sup>1</sup>—and survivors of severe infections receive longer-lasting immunity



**Figure 2 | Profile likelihood of the duration of immunity,  $1/\epsilon$ , in the SIRS model for the data from Dacca district.** Dacca (now spelled Dhaka) is the only district for which comparable analyses have been performed<sup>13,27</sup>. To compute the profile likelihood,  $\epsilon$  is fixed and the likelihood is maximized over the remaining parameters. The vertical dashed lines embrace the approximate 95% confidence interval (2–12 weeks).

( $1.5 \pm 0.7$  yr), consistent with experimental evidence<sup>14,15,19</sup>. Although in some districts the evidence strongly favours the two-path model, the data taken as a whole are equivocal on the subject of which model is better (see Supplementary Table 2). The implication is that it is the short-term immunity and high asymptomatic ratio predicted by both models, and not a large fraction of silent shedders, that accounts for the dramatically improved description afforded by our models.

A second robust prediction of both models is that, in typical epidemics, relatively few cases are due to the environmental reservoir. More precisely, though our models cannot speak to the exact route of transmission (that is, food-borne, water-borne, fomites, etc.), our results clearly indicate the relative importance of the positive feedback associated with the human-associated source of infection over the human-independent environmental source of infection. Nevertheless, the seasonally averaged basic reproductive number,  $R_0$ , for human-associated transmission is estimated to be quite low (SIRS model,  $1.5 \pm 0.2$ ; two-path model,  $1.5 \pm 0.2$ ). This contrasts sharply with the much higher values of  $R_0$  recently proposed<sup>20</sup> and suggests, paradoxically, that for *V. cholerae*, in this the region where cholera is most persistent and outbreaks most frequent, humans are potentially a marginal habitat: the persistence of the disease may be largely due to the environmental reservoir which provides a small extrinsic force of infection. The strength of this environmental force of infection varies geographically in a pattern (Supplementary Fig. 2) that mirrors the earlier suggestion that the coastal regions of Bangladesh are the native habitat of classical cholera<sup>21</sup>. Both the SIRS and two-path models make the simplifying assumption that the force of infection due to the environmental reservoir has negligible seasonal fluctuation. To relax this assumption<sup>22</sup>, we fit a third model which extends the SIRS model but allows for seasonality in the environmental reservoir. This model explains the data significantly better in many districts, but our conclusions about low  $R_0$ , rapid loss of immunity and high prevalence of inapparent infection remain unchanged (see Supplementary Information).

A central challenge in cholera epidemiology is to explain how the outbreaks can be so explosive at the outset and yet self-limiting to the point that sizable epidemics can recur twice annually. Our explanation contrasts sharply with those proposed earlier<sup>13,17</sup>. Simulations of the various models (Fig. 3) make plain the contrasts between the explanations. In the previous view, infection-derived immunity tends to wane over the scale of years, the explosiveness of outbreaks is controlled by spatial effects, and it is the seasonal drop in transmissibility (presumably related to environmental drivers) that stops the epidemics. Because in the previous view the ratio of deaths to infections is comparatively modest, the susceptible population is depleted only slightly by each epidemic and is continually replenished by births and the waning of immunity. In the new view, epidemics are again triggered by rising seasonal transmissibility and the presence of a reservoir, but the high asymptomatic ratio means that many more individuals are exposed. The vast majority of these receive short-term protection from infection so that depletion of the susceptible pool brings the epidemic to a halt. As immunity wanes, on the timescale of weeks to months, the susceptible pool is replenished, setting the stage for the next outbreak. In the new view, severe cases play little role in shaping the dynamics.

It has been proposed that the seasonality of cholera in this region may be due to the interaction of free-living vibrio with lytic bacteriophage in surface waters<sup>23–25</sup>. According to this hypothesis, as an epidemic progresses, the density of phage in the environment increases and survivorship of vibrio consequently declines, which in turn leads to attenuation of transmissibility. To determine whether such an effect might explain the data, we formulated a fourth model, which extends the SIRS model by allowing for phage to build up in the environment (as infected individuals shed pathogen into surface waters) and ultimately interrupt transmission (Fig. 1c and Supplementary Equations). Using the same methods, we estimated

the strength and timescale of this effect; the results are presented in the Supplementary Information. In brief, the main conclusions described above are robust to the inclusion of the environmental phage effect: the estimated duration of immunity remains short ( $9.3 \pm 8.3$  weeks), fatality low ( $0.004 \pm 0.002$ , compare the corresponding value for the SIRS model) and  $R_0$  small ( $1.6 \pm 0.3$ ). Overall, the historical mortality data are not strong evidence in favour of the environmental-phage hypothesis. Furthermore, a critical implication of the environmental-phage hypothesis is that the phage cycles must lag behind cholera cycles, but not too far: the initial drop in cholera cases is predicted to be roughly coincident with high levels of ambient phage. In the one instance where concurrent data on cholera cases and ambient phage densities are available, however, the phage appear to lag approximately 180 degrees behind cholera cases<sup>24</sup>. Thus cholera cases begin to fall while phage concentrations remain small, and cholera cases begin to climb again while phage densities remain high. This is further, independent, evidence against the notion that phage–vibrio interaction in the environment is responsible for cholera seasonality.

The models presented here make several testable predictions. In particular, our results suggest that population surveys using sensitive

techniques to detect the presence of *V. cholerae*, cholera-specific bacteriophage and/or coproantibodies in stools of healthy and mildly symptomatic individuals should reveal a much higher prevalence of inapparent infection than is currently understood to be the case. Such inapparent infections are predicted to be associated with reduced susceptibility to infection. Studies that centre on cholera patients, on the other hand, because they focus on the tail of the distribution of disease severity, can be expected to yield severely biased results when extrapolated to the population level. Further implications of our results for public health are discussed in the Supplementary Information.

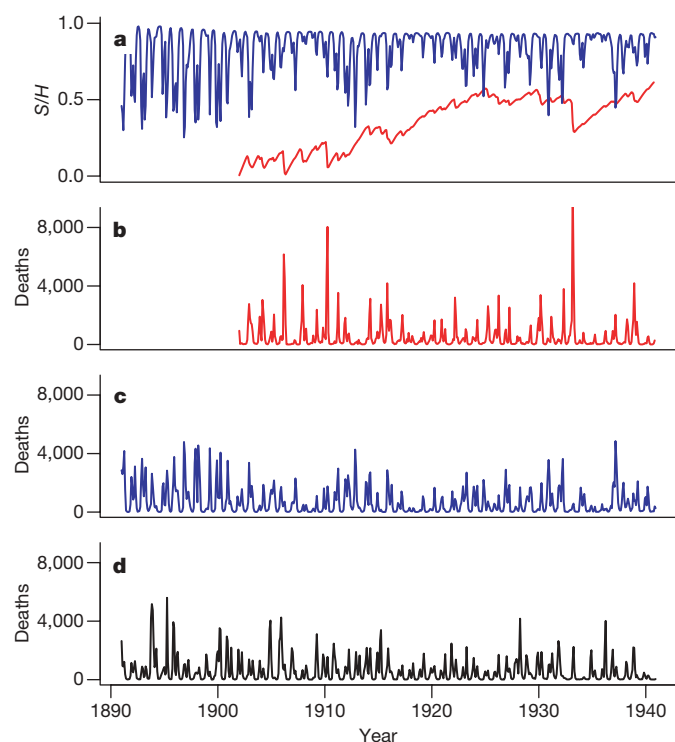
In all models of epidemiological dynamics, critical quantities include the fraction of individuals susceptible to infection, immune from infection and asymptomatically infected. Yet it is almost never the case that available data include direct measurements of these quantities. Moreover, it is frequently true that many asymptomatic—or merely unrecorded—infections exist for each case that is recorded. Our results demonstrate that conclusions about the biological mechanisms underlying a pattern of disease outbreaks can depend sensitively on the prevalence of inapparent infections. New modelling approaches that allow indirect inference in the face of hidden variables and incomplete information are therefore likely to revise our understanding of the ecology of many infectious diseases.

## METHODS SUMMARY

All models were formulated as stochastic differential equations, which were integrated using the Euler–Maruyama algorithm. Details of the equations and simulation methods are given in the Methods. The iterated filtering algorithm<sup>7</sup> was used to fit the models to the data. This algorithm is described in detail in the Supplementary Methods and is implemented as part of an open-source package, pomp, within the R statistical computing environment<sup>26</sup>. All codes will be made available by the authors on request.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 25 March; accepted 13 May 2008.



**Figure 3 | Typical model simulations versus data for the district of Dacca.**

**a**, Simulated susceptible fraction ( $S/H$ ) from our SIRS model with seasonal reservoir (blue) and from the time-series SIRS (TSIRS) model of Koelle *et al.*<sup>13,17</sup> (red). The estimated susceptible fraction of the population is much lower under the TSIRS model, and grows with time. By contrast, under the models introduced here, rapid waning of immunity allows the susceptible fraction to return seasonally to high levels. The predicted high prevalence of inapparent infection implies that the susceptible pool is rapidly depleted during an outbreak. **b**, Simulated monthly deaths from the TSIRS model. Despite the low susceptible fraction and that model's non-mass-action transmission assumption, the relatively low prevalence of inapparent infection leads to overly explosive outbreaks and overly deep crashes. **c**, Simulated monthly deaths from the SIRS model with seasonal reservoir. **d**, Cholera death data from Dacca district. Parameter estimates for the seasonal-reservoir SIRS model are reported in the Supplementary Information; parameters for the TSIRS model are from K. Koelle (personal communication). The semi-parametric regression used to fit the TSIRS model is initialized using the first 11 yr of data; no simulated data are therefore generated over that interval.

1. Kaper, J. B., Morris, J. G. & Levine, M. M. Cholera. *Clin. Microbiol. Rev.* **8**, 48–86 (1995).
2. Feachem, R. G. Environmental aspects of cholera epidemiology. III. Transmission and control. *Trop. Dis. Bull.* **79**, 1–47 (1982).
3. Van De Linde, P. A. M. & Forbes, G. I. Observations on the spread of cholera in Hong Kong, 1961–63. *Bull. World Health Organ.* **32**, 515–530 (1965).
4. McCormack, W. M., Islam, M. S., Fahimuddin, M. & Mosley, W. H. A community study of inapparent cholera infections. *Am. J. Epidemiol.* **89**, 658–664 (1969).
5. Glass, R. I. & Black, R. E. in *Cholera* (eds Barua, D. & Greenough, W. B. III) 129–154 (Plenum Medical Book Co., New York, 1992).
6. Islam, M. S., Drasar, B. S. & Sack, R. B. The aquatic environment as a reservoir of *Vibrio cholerae*: a review. *J. Diarrhoeal Dis. Res.* **11**, 197–206 (1993).
7. Ionides, E. L., Bretó, C. & King, A. A. Inference for nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **103**, 18438–18443 (2006).
8. Sack, D. A., Sack, R. B., Nair, G. B. & Siddique, A. K. Cholera. *Lancet* **363**, 223–233 (2004).
9. Sanitary Commissioner for Bengal Reports and Bengal Public Health Reports. (Bengal Secretariat Press, Calcutta and Bengal Government Press, Alipore, 1891–, 1942).
10. Glass, R. I., Lee, J. V., Huq, M. I., Hossain, K. M. & Khan, M. R. Phage types of *Vibrio cholerae* O1 biotype El Tor isolated from patients and family contacts in Bangladesh: epidemiologic implications. *J. Infect. Dis.* **148**, 998–1004 (1983).
11. Sack, R. B. *et al.* A 4-year study of the epidemiology of *Vibrio cholerae* in four rural areas of Bangladesh. *J. Infect. Dis.* **187**, 96–101 (2003).
12. Huq, A. *et al.* Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl. Environ. Microbiol.* **71**, 4645–4654 (2005).
13. Koelle, K. & Pascual, M. Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera. *Am. Nat.* **163**, 901–913 (2004).
14. Levine, M. M. *et al.* in *Acute Enteric Infections in Children: New Prospects for Treatment and Prevention* (eds Holme, T., Holmgren, J., Merson, M. H. & Mollby, R.) 443–459 (Elsevier/North-Holland Biomedical Press, Amsterdam, 1981).
15. Cash, R. A. *et al.* Response of man to infection with *Vibrio cholerae*. II. Protection from illness afforded by previous disease and vaccine. *J. Infect. Dis.* **130**, 325–333 (1974).
16. Glass, R. I. *et al.* Endemic cholera in rural Bangladesh, 1966–1980. *Am. J. Epidemiol.* **116**, 959–970 (1982).
17. Koelle, K., Rodo, X., Pascual, M., Yunus, M. & Mostafa, G. Refractory periods and climate forcing in cholera dynamics. *Nature* **436**, 696–700 (2005).

18. World Health Organization. *Wkly Epidemiol. Rec.* (<http://www.who.int/wer>) (1950–, 1965).
19. Cash, R. A. *et al.* Response of man to infection with *Vibrio cholerae*. I. Clinical, serologic, and bacteriologic responses to a known inoculum. *J. Infect. Dis.* **129**, 45–52 (1974).
20. Hartley, D. M., Morris, J. G. & Smith, D. L. Hyperinfectivity: a critical element in the ability of *V. cholerae* to cause epidemics? *PLoS Med.* **3**, 63–69 (2006).
21. Siddique, A. K. *et al.* Survival of classic cholera in Bangladesh. *Lancet* **337**, 1125–1127 (1991).
22. Bouma, M. J. & Pascual, M. Seasonal and interannual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologia* **460**, 147–156 (2001).
23. Faruque, S. M. *et al.* Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. *Proc. Natl Acad. Sci. USA* **102**, 6119–6124 (2005).
24. Faruque, S. M. *et al.* Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. *Proc. Natl Acad. Sci. USA* **102**, 1702–1707 (2005).
25. Jensen, M. A., Faruque, S. M., Mekalanos, J. J. & Levin, B. R. Modeling the role of bacteriophage in the control of cholera outbreaks. *Proc. Natl Acad. Sci. USA* **103**, 4652–4657 (2006).
26. R Development Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2007).
27. Rodo, X., Pascual, M., Fuchs, G. & Faruque, A. S. G. ENSO and cholera: a nonstationary link related to climate change? *Proc. Natl Acad. Sci. USA* **99**, 12901–12906 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge discussions with K. Koelle, C. Bretó and A.P. Dobson. This work was funded by the joint National Science Foundation–National Institutes of Health Ecology of Infectious Diseases Program (NSF grants nos 0545276 and 0430120) and the National Oceanic and Atmospheric Administration (Oceans and Health, grant no. NAO40AR460019).

**Author Contributions** A.A.K. formulated and implemented the models, implemented the fitting algorithm, performed the analyses, and drafted and revised the text. E.L.I. provided input on the models and statistical analyses, and drafted much of the Supplementary Information. M.P. provided input on the models and commented on the text. M.J.B. extracted and assembled the data and provided input on the models and on the historical and clinical aspects of cholera.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.A.K. ([kingaa@umich.edu](mailto:kingaa@umich.edu)).

## METHODS

**Inference for dynamical systems.** A stochastic dynamical system is generated by a suitable function  $f(x, s, t, \theta, W)$ . Here,  $W$  is a stochastic quantity that is drawn independently each time the function is evaluated and  $\theta$  is a vector of parameters. Thus, for a sequence of times  $t_0 \leq t_1 \leq \dots \leq t_N$ , the recursion

$$X(t_n) = f(X(t_{n-1}), t_{n-1}, t_n, \theta, W),$$

together with the initial value  $X(t_0)$ , specifies a Markov process. The unspecified nature of  $f(\cdot)$  reflects the idea that the evolution of the system could correspond to an essentially arbitrary stochastic simulation algorithm. In the context of this paper,  $X(t)$  is a vector of counts in each disease class and  $f(\cdot)$  corresponds to a (numerical) solution of any of the stochastic differential equation systems in the Supplementary Equations. The data consist of observations  $y_1, \dots, y_n$  at times  $t_1, \dots, t_n$ . We suppose that, conditional on  $\{X(t_k) = x_k\}_{k=1}^n, y_n$  is drawn from a density  $g(y_n | x_n, t_n, \theta)$ .

In the Supplementary Methods, we present a pseudocode description of an iterated filtering methodology (maximization by iterated filtering, MIF) which has recently been shown to enable likelihood-based inference for this framework<sup>7</sup>. This procedure allows inference for rather general partially observed, stochastic, multivariate, continuous-time dynamical systems. The MIF algorithm also accommodates measurement error, non-stationarity, irregular observation times and the mechanistic inclusion of covariates. It has been implemented in the R statistical language<sup>26</sup> as part of the package *pomp*. Note that the algorithmic parameters play a role in the computational efficiency of the method but do not affect the scientific conclusions once likelihood maximization has been confirmed by diagnostic plots<sup>7</sup>. The package and all other codes will be made available by the first author on request.

**Fitting the Bengal data.** Our statistical inference technique is based on simulation of the stochastic dynamical system. To simulate the stochastic dynamical systems, we used the Euler–Maruyama numerical scheme<sup>28</sup> with a timestep of 0.05 months. We fit the parameters of each model to the data from each district using the algorithm given in the Supplementary Methods. We used  $J = 10,000$  particles, a fixed lag  $L = 60$  and  $M = 80$  iterations with a cooling factor,  $a = 0.95$ . Using the same algorithm, we then constructed profile likelihoods against the parameters  $\varepsilon$  and  $c$ . This analysis revealed the existence of weak non-identifiabilities among the parameters  $c, \gamma, m, \varepsilon$  and  $\rho$ . Despite this, certain combinations of parameters, such as  $R_0 = c(\beta_{\text{seas}}(t))/(\gamma + \delta + m)$  and the case fatality  $m/(\gamma + \delta + m)$ , are well-identified. Moreover, although estimates of the durations of short-term immunity vary considerably among districts, it is the case that, for all districts, the duration of immunity is predicted to be far shorter than that derived from severe infections induced experimentally<sup>13–17</sup>. Construction of the profile likelihoods also facilitated the location of the global maximum likelihood estimates, which, after further refinement using the algorithm above with 30,000 particles, are reported in the Supplementary Tables.

**Review of likelihood-based inference.** We give a review of likelihood-based inference in the context of this paper. This review focuses on practical considerations; we refer the reader to Casella & Berger<sup>29</sup> or Rice<sup>30</sup> for a complete introduction, and to Barndorff-Nielsen & Cox<sup>31</sup> for a more advanced treatment. For a discussion of the role of likelihood, profile likelihood ratio tests applied to stochastic models for biological systems, we direct the reader to Hilborn & Mangel<sup>32</sup>. Reasons to use likelihood as a basis for inference include statistical efficiency (that is, making the strongest conclusions available from limited data), objectivity (that is, providing a single criterion applicable to a wide range of models and data) and the availability of useful techniques for comparing rival hypotheses.

For time-series data, written as  $x_{1:T} = x_1, \dots, x_T$ , a stochastic model consists of a density function  $f(x_{1:T} | \theta)$  depending on a vector  $\theta$  of unknown parameters. The likelihood function arises from considering the density as a function of  $\theta$ , treating the data as fixed. Thus the log likelihood is defined to be

$$\log \mathcal{L}(\theta) = \log f(x_{1:T} | \theta),$$

where  $\log$  is the natural logarithm. Rival hypotheses, labelled  $H_0$  and  $H_1$ , concerning the appropriate structure for a stochastic model can be formalized in terms of whether  $\theta$  takes a value in  $\Theta_0$  or  $\Theta_1$  respectively, where  $\Theta_0$  and  $\Theta_1$  are subsets of the set  $\Theta$  of all possible parameter vectors. We refer to  $H_0$  and  $H_1$  as hypotheses or models interchangeably.

**Comparison of nested models.** Suppose that  $H_0$  and  $H_1$  are nested, meaning that  $\Theta_0$  is a subset of  $\Theta_1$ . In practical terms, this is taken to mean that  $\Theta_0$  is described by  $d_0$  freely varying parameters, whereas  $\Theta_1$  is described by these  $d_0$  parameters together with a number  $d_1 - d_0$  of additional parameters. Thus  $\Theta_0$  can be written in terms of a vector of length  $d_0$  and  $\Theta_1$  in terms of a vector of length  $d_1$ . The

maximum likelihood parameter estimate for model  $H_0$  is denoted by  $\hat{\theta}_0$ , and corresponds to the parameter value giving rise to the largest value of  $\log \mathcal{L}(\theta)$  among all values of  $\theta$  in  $\Theta_0$ . A corresponding definition applies to  $\hat{\theta}_1$ . Because  $\Theta_1$  includes  $\Theta_0$ , it follows that  $\log \mathcal{L}(\hat{\theta}_1) \geq \log \mathcal{L}(\hat{\theta}_0)$ . Under standard regularity conditions, twice the improvement in the log likelihood that occurs when moving from the simpler model  $H_0$  to the more complex model  $H_1$  can be compared with a  $\chi^2$  random variable on  $d_1 - d_0$  degrees of freedom. Thus, Supplementary Table 1 presents a  $P$ -value of the form

$$P = \text{Prob} \left[ \chi_{d_1 - d_0}^2 > 2 \left\{ \log \mathcal{L}(\hat{\theta}_1) - \log \mathcal{L}(\hat{\theta}_0) \right\} \right]. \quad (1)$$

The  $P$ -value given by equation (1) is an approximation. More precise  $P$ -values can in principle be obtained by simulation, but this  $\chi^2$  approximation is a widely used technique that has been found to be reliable for many practical purposes<sup>33</sup>. It is both conceptually and operationally simpler to calculate initially the standard  $\chi^2$  approximation in equation (1); if this gives rise to an unequivocal answer, there is little reason to present a simulation study as additional evidence.

Modifications to equation (1) are required when the models under consideration are nested in a non-standard way<sup>34</sup>. This arises in the context of the analysis presented in Supplementary Table 1, owing to the non-standard nesting of the two-path model within the SIRS model. Specifically, the two-path model reduces to the SIRS model when  $c = 1$  at which point  $\rho$  is unidentified. Anisimova *et al.*<sup>35</sup> investigated similar situations and demonstrated that the chi-squared approximation to the likelihood ratio test is conservative in such cases. In the caption of Supplementary Table 1, we report the  $P$ -value associated with such a test on three degrees of freedom:  $\rho, c$  and the initial condition  $Y_0$ .

Non-nested models may still be compared by their likelihood, for example by the use of Akaike's information criterion<sup>36,37</sup>, but formal testing of hypotheses becomes more awkward. Thus, in Supplementary Table 1, we contrast the best-fitting SIRS model with the nested SIRS model constrained to match the duration of immunity found by Koelle & Pascual<sup>13</sup>. Another consideration motivating this comparison is that the nested SIRS model has a similar likelihood to the model of Koelle & Pascual<sup>13</sup>, while having fewer parameters to estimate. Additionally, this approach avoids the issue that the comparison with previously published models might be unfair because the procedure used to fit the models in this paper may be more effective than that used by previous research.

**Profile likelihood.** The profile likelihood is a description of the weight of evidence about the value of a single component of the vector of unknown parameters. Setting  $\theta = (\theta_1, \dots, \theta_d)$ , the log likelihood can be written as  $\log \mathcal{L}(\theta_1, \dots, \theta_d)$  and the profile log likelihood of  $\theta_1$  is

$$\log \mathcal{L}_{\text{profile}}(\theta_1) = \max_{\theta_2, \dots, \theta_d} \log \mathcal{L}(\theta_1, \dots, \theta_d).$$

From equation (1), a 95% confidence region for  $\theta_1$  is given by the range of values of  $\theta_1$  for which

$$2 \left( \log \mathcal{L}(\hat{\theta}) - \log \mathcal{L}_{\text{profile}}(\theta_1) \right) < c,$$

where  $c$  is defined by  $\text{Prob}[\chi_1^2 < c] = 0.95$ . This is how the confidence interval in Fig. 2 was constructed.

28. Kloeden, P. E. & Platen, E. *Numerical Solution of Stochastic Differential Equations* 3rd edn (Springer, New York, 1999).
29. Casella, G. & Berger, R. L. *Statistical Inference* (Wadsworth, Pacific Grove, 1990).
30. Rice, J. A. *Mathematical Statistics and Data Analysis* (Wadsworth, Belmont, California, 1988).
31. Barndorff-Nielsen, O. E. & Cox, D. R. *Inference and Asymptotics* (Chapman and Hall, London, 1994).
32. Hilborn, R. & Mangel, M. *The Ecological Detective* (Princeton University Press, Princeton, New Jersey, 1997).
33. McCullagh, P. & Nelder, J. A. *Generalized Linear Models* 2nd edn (Chapman and Hall, London, 1989).
34. Self, S. G. & Liang, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* **82**, 605–610 (1987).
35. Anisimova, M., Bielawski, J. P. & Yang, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**, 1585–1592 (2001).
36. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
37. Burnham, K. P. & Anderson, D. R. *Model Selection and Inference: A Practical Information-Theoretic Approach* 2nd edn (Springer, New York, 2002).

# Inapparent infections and cholera dynamics

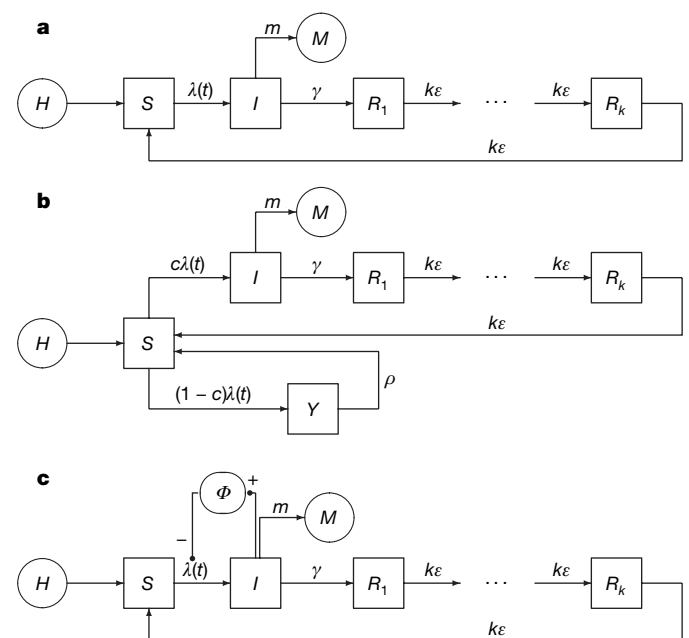
Aaron A. King<sup>1,2</sup>, Edward L. Ionides<sup>3</sup>, Mercedes Pascual<sup>1,4</sup> & Menno J. Bouma<sup>5</sup>

In many infectious diseases, an unknown fraction of infections produce symptoms mild enough to go unrecorded, a fact that can seriously compromise the interpretation of epidemiological records. This is true for cholera, a pandemic bacterial disease, where estimates of the ratio of asymptomatic to symptomatic infections have ranged from 3 to 100 (refs 1–5). In the absence of direct evidence, understanding of fundamental aspects of cholera transmission, immunology and control has been based on assumptions about this ratio and about the immunological consequences of inapparent infections. Here we show that a model incorporating high asymptomatic ratio and rapidly waning immunity, with infection both from human and environmental sources, explains 50 yr of mortality data from 26 districts of Bengal, the pathogen's endemic home. We find that the asymptomatic ratio in cholera is far higher than had been previously supposed and that the immunity derived from mild infections wanes much more rapidly than earlier analyses have indicated. We find, too, that the environmental reservoir<sup>5,6</sup> (free-living pathogen) is directly responsible for relatively few infections but that it may be critical to the disease's endemicity. Our results demonstrate that inapparent infections can hold the key to interpreting the patterns of disease outbreaks. New statistical methods<sup>7</sup>, which allow rigorous maximum likelihood inference based on dynamical models incorporating multiple sources and outcomes of infection, seasonality, process noise, hidden variables and measurement error, make it possible to test more precise hypotheses and obtain unexpected results. Our experience suggests that the confrontation of time-series data with mechanistic models is likely to revise our understanding of the ecology of many infectious diseases.

Cholera is a diarrhoeal disease caused by enteric infection with the bacterium *Vibrio cholerae*. Six of the seven cholera pandemics that have swept the globe since 1817 originated in the low-lying, densely populated regions north of the Bay of Bengal, where the disease is endemic. Although much attention has been focused on cholera<sup>1,8</sup>, unsolved puzzles remain about its mode of transmission and the role of host immunity in its dynamics. This is largely because, in regions where cholera is endemic, most cholera cases are mild or asymptomatic but the true extent of asymptomatic infection has been difficult to assess. Estimates of the ratio of asymptomatic to symptomatic cases vary greatly, and the importance of inapparent infections in the dynamics of cholera outbreaks is unknown. To determine what role is played by inapparent infections, we used an approach that allows indirect inference about unobserved variables.

A remarkably rich data set on the pattern of cholera epidemics exists in the form of mortality records kept by the sanitary commissioners of the former British East Indian province of Bengal<sup>9</sup>. The data consist of monthly cholera death counts in each of 26 districts over the period 1891–1940 (Supplementary Fig. 1). To analyse these data, we formulated a series of models incorporating known or hypothesized mechanisms of transmission and immunity. A

parsimonious model for cholera dynamics is of susceptible–infectious–recovered–susceptible (SIRS) form (Fig. 1a). A novel feature of this model is that it incorporates both transmission tied to human prevalence (using a traditional mass-action term) and transmission from an environmental reservoir (where the pathogen is commonly living in aquatic environments)<sup>5,6,10–12</sup>. This model is a



**Figure 1 | The mechanistic models used.** **a**, SIRS model; **b**, two-path model; **c**, environmental-phage model. Births, related to the total population size  $H$ , are assumed to feed the pool of susceptibles,  $S$ . Individuals are susceptible to infection when born. Exposure to the pathogen occurs at time-dependent rate  $\lambda(t)$ .  $c$  is the probability that an exposure leads to a contagious infection (class  $I$ ). Note that when  $c = 1$  and  $\rho = \infty$ , the two-path model (**b**) reduces to the SIRS model (**a**); when  $c < 1$ , some exposures result in short-term immunity (class  $Y$ ). Infected individuals die at an excess rate  $m$  and recover at a rate  $\gamma$ ; the time an individual spends within the  $I$  class is exponentially distributed. We assume that an individual remains immune to reinfection for a duration gamma-distributed with mean  $1/\epsilon$  and variance  $1/k\epsilon^2$ . Once immunity has waned, an individual re-enters the susceptible pool ( $S$ ). The measured variable is monthly deaths,  $M$ . The mean duration of short-term immunity is  $1/\rho$ . Individuals in each class are subject to constant background mortality at rate  $0.02 \text{ yr}^{-1}$ . The force of infection,  $\lambda(t)$ , includes terms for environmental and human sources of infection and is assumed to vary seasonally. Because the seasonality of cholera dynamics in Bengal is complex, we used a semi-mechanistic approach: transmission was modelled by a flexible periodic function of time. In the environmental-phage model (**c**), as infected hosts shed pathogen, phage  $\Phi$  builds up in the environment and reduces transmissibility. The equations specifying these models are given in the Supplementary Equations.

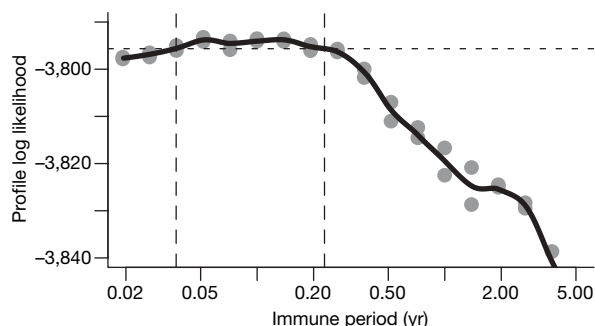
<sup>1</sup>Department of Ecology and Evolutionary Biology, <sup>2</sup>Department of Mathematics, <sup>3</sup>Department of Statistics, University of Michigan, Ann Arbor, Michigan 48109, USA. <sup>4</sup>Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA. <sup>5</sup>Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, University of London, London WC1E 7HT, UK.

partially observed, nonlinear, stochastic dynamical system in continuous time and as such is not amenable to analysis by standard statistical techniques. Inference for such systems has recently been facilitated by a new likelihood maximization procedure<sup>7</sup> which allows for measurement error, non-stationarity, irregular sampling intervals and the mechanistic inclusion of covariates (see Methods). We fitted the SIRS model to each district's data, assessed its explanatory power using log likelihood and the Akaike information criterion (AIC<sub>c</sub>) and compared it with a semi-mechanistic time-series model for cholera<sup>13</sup> and two non-mechanistic time-series models.

The SIRS model explained the data dramatically better than the previous best fit<sup>13</sup> ( $\Delta\text{AIC}_c > 58$ ; Supplementary Table 1). Parameter estimates reveal some surprises. A prediction of the model, robustly consistent across districts, is that immunity must wane on a timescale of weeks to months (Fig. 2 and Supplementary Table 3). This is in stark contrast to the widely held belief that infection-derived immunity to cholera wanes on a timescale of 3–10 yr (refs 13–17). The model also robustly predicts low case fatality ( $0.004 \pm 0.002$ , mean  $\pm$  s.d. across 26 districts). Because fatality among hospitalized cases was historically in excess of 50%<sup>18</sup>, this implies a very large asymptomatic ratio. The implication is that most exposures result not in severe cholera, but in mild or asymptomatic infection, the immunological memory of which is short-lived. This prediction of a high asymptomatic ratio is consistent with the only intensive field studies of inapparent infections of which we are aware<sup>3,4</sup>. However, because in this model all infected individuals are equally infectious, the prediction is that the vast majority of infectives are 'silent shedders'—infectious but without symptoms. The evidence for such a conclusion is mixed but weak at best<sup>2</sup>. Moreover, because in this model all infections are alike, it is unclear whether the improved dynamical description afforded by the SIRS model over the previous best fit is due to the higher proportion of asymptomatic infections or to the shorter duration of immunity.

To tease apart the immunological consequences of exposure to the pathogen from the degree of infectiousness induced, and to allow for heterogeneity in infectiousness, we formulated a second mechanistic model in which exposed individuals can follow an alternate pathway, deriving immunity to infection while shedding a negligible quantity of vibrio and suffering a negligible disease-induced mortality (Fig. 1b). In essence, this model allows for the possibility that exposure to the pathogen effectively vaccinates against reinfection. We leave it to the data to determine how frequently such natural vaccination occurs and how long the resulting immunity lasts.

Under the two-path model, the timescale of short-term immunity ( $9.9 \pm 4.7$  weeks, mean  $\pm$  s.d. across 26 districts) agrees with that of the SIRS model. The estimated case fatality under this model is  $0.34 \pm 0.21$ —consistent with conditions during this period<sup>1</sup>—and survivors of severe infections receive longer-lasting immunity



**Figure 2 | Profile likelihood of the duration of immunity,  $1/\epsilon$ , in the SIRS model for the data from Dacca district.** Dacca (now spelled Dhaka) is the only district for which comparable analyses have been performed<sup>13,27</sup>. To compute the profile likelihood,  $\epsilon$  is fixed and the likelihood is maximized over the remaining parameters. The vertical dashed lines embrace the approximate 95% confidence interval (2–12 weeks).

( $1.5 \pm 0.7$  yr), consistent with experimental evidence<sup>14,15,19</sup>. Although in some districts the evidence strongly favours the two-path model, the data taken as a whole are equivocal on the subject of which model is better (see Supplementary Table 2). The implication is that it is the short-term immunity and high asymptomatic ratio predicted by both models, and not a large fraction of silent shedders, that accounts for the dramatically improved description afforded by our models.

A second robust prediction of both models is that, in typical epidemics, relatively few cases are due to the environmental reservoir. More precisely, though our models cannot speak to the exact route of transmission (that is, food-borne, water-borne, fomites, etc.), our results clearly indicate the relative importance of the positive feedback associated with the human-associated source of infection over the human-independent environmental source of infection. Nevertheless, the seasonally averaged basic reproductive number,  $R_0$ , for human-associated transmission is estimated to be quite low (SIRS model,  $1.5 \pm 0.2$ ; two-path model,  $1.5 \pm 0.2$ ). This contrasts sharply with the much higher values of  $R_0$  recently proposed<sup>20</sup> and suggests, paradoxically, that for *V. cholerae*, in this the region where cholera is most persistent and outbreaks most frequent, humans are potentially a marginal habitat: the persistence of the disease may be largely due to the environmental reservoir which provides a small extrinsic force of infection. The strength of this environmental force of infection varies geographically in a pattern (Supplementary Fig. 2) that mirrors the earlier suggestion that the coastal regions of Bangladesh are the native habitat of classical cholera<sup>21</sup>. Both the SIRS and two-path models make the simplifying assumption that the force of infection due to the environmental reservoir has negligible seasonal fluctuation. To relax this assumption<sup>22</sup>, we fit a third model which extends the SIRS model but allows for seasonality in the environmental reservoir. This model explains the data significantly better in many districts, but our conclusions about low  $R_0$ , rapid loss of immunity and high prevalence of inapparent infection remain unchanged (see Supplementary Information).

A central challenge in cholera epidemiology is to explain how the outbreaks can be so explosive at the outset and yet self-limiting to the point that sizable epidemics can recur twice annually. Our explanation contrasts sharply with those proposed earlier<sup>13,17</sup>. Simulations of the various models (Fig. 3) make plain the contrasts between the explanations. In the previous view, infection-derived immunity tends to wane over the scale of years, the explosiveness of outbreaks is controlled by spatial effects, and it is the seasonal drop in transmissibility (presumably related to environmental drivers) that stops the epidemics. Because in the previous view the ratio of deaths to infections is comparatively modest, the susceptible population is depleted only slightly by each epidemic and is continually replenished by births and the waning of immunity. In the new view, epidemics are again triggered by rising seasonal transmissibility and the presence of a reservoir, but the high asymptomatic ratio means that many more individuals are exposed. The vast majority of these receive short-term protection from infection so that depletion of the susceptible pool brings the epidemic to a halt. As immunity wanes, on the timescale of weeks to months, the susceptible pool is replenished, setting the stage for the next outbreak. In the new view, severe cases play little role in shaping the dynamics.

It has been proposed that the seasonality of cholera in this region may be due to the interaction of free-living vibrio with lytic bacteriophage in surface waters<sup>23–25</sup>. According to this hypothesis, as an epidemic progresses, the density of phage in the environment increases and survivorship of vibrio consequently declines, which in turn leads to attenuation of transmissibility. To determine whether such an effect might explain the data, we formulated a fourth model, which extends the SIRS model by allowing for phage to build up in the environment (as infected individuals shed pathogen into surface waters) and ultimately interrupt transmission (Fig. 1c and Supplementary Equations). Using the same methods, we estimated

the strength and timescale of this effect; the results are presented in the Supplementary Information. In brief, the main conclusions described above are robust to the inclusion of the environmental phage effect: the estimated duration of immunity remains short ( $9.3 \pm 8.3$  weeks), fatality low ( $0.004 \pm 0.002$ , compare the corresponding value for the SIRS model) and  $R_0$  small ( $1.6 \pm 0.3$ ). Overall, the historical mortality data are not strong evidence in favour of the environmental-phage hypothesis. Furthermore, a critical implication of the environmental-phage hypothesis is that the phage cycles must lag behind cholera cycles, but not too far: the initial drop in cholera cases is predicted to be roughly coincident with high levels of ambient phage. In the one instance where concurrent data on cholera cases and ambient phage densities are available, however, the phage appear to lag approximately 180 degrees behind cholera cases<sup>24</sup>. Thus cholera cases begin to fall while phage concentrations remain small, and cholera cases begin to climb again while phage densities remain high. This is further, independent, evidence against the notion that phage–vibrio interaction in the environment is responsible for cholera seasonality.

The models presented here make several testable predictions. In particular, our results suggest that population surveys using sensitive

techniques to detect the presence of *V. cholerae*, cholera-specific bacteriophage and/or coproantibodies in stools of healthy and mildly symptomatic individuals should reveal a much higher prevalence of inapparent infection than is currently understood to be the case. Such inapparent infections are predicted to be associated with reduced susceptibility to infection. Studies that centre on cholera patients, on the other hand, because they focus on the tail of the distribution of disease severity, can be expected to yield severely biased results when extrapolated to the population level. Further implications of our results for public health are discussed in the Supplementary Information.

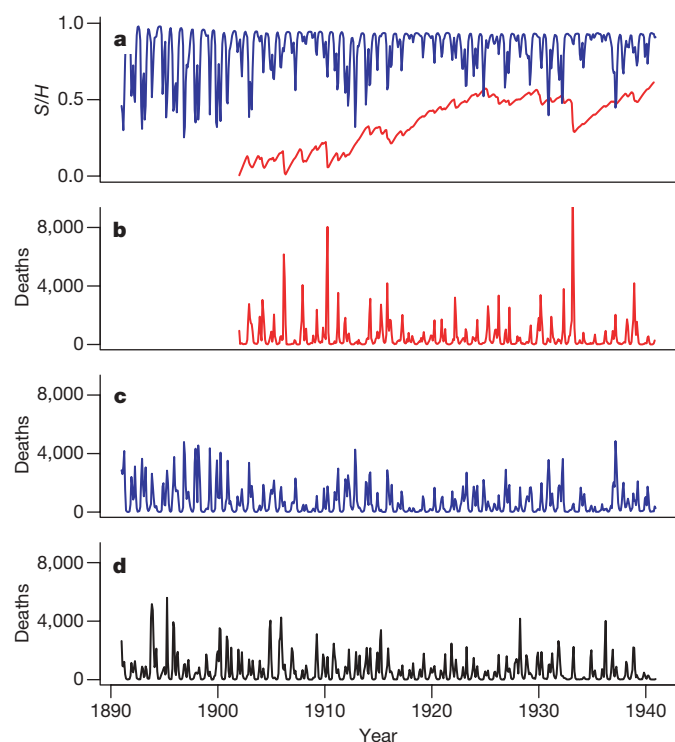
In all models of epidemiological dynamics, critical quantities include the fraction of individuals susceptible to infection, immune from infection and asymptomatically infected. Yet it is almost never the case that available data include direct measurements of these quantities. Moreover, it is frequently true that many asymptomatic—or merely unrecorded—infections exist for each case that is recorded. Our results demonstrate that conclusions about the biological mechanisms underlying a pattern of disease outbreaks can depend sensitively on the prevalence of inapparent infections. New modelling approaches that allow indirect inference in the face of hidden variables and incomplete information are therefore likely to revise our understanding of the ecology of many infectious diseases.

## METHODS SUMMARY

All models were formulated as stochastic differential equations, which were integrated using the Euler–Maruyama algorithm. Details of the equations and simulation methods are given in the Methods. The iterated filtering algorithm<sup>7</sup> was used to fit the models to the data. This algorithm is described in detail in the Supplementary Methods and is implemented as part of an open-source package, pomp, within the R statistical computing environment<sup>26</sup>. All codes will be made available by the authors on request.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 25 March; accepted 13 May 2008.



**Figure 3 | Typical model simulations versus data for the district of Dacca.**

**a**, Simulated susceptible fraction ( $S/H$ ) from our SIRS model with seasonal reservoir (blue) and from the time-series SIRS (TSIRS) model of Koelle *et al.*<sup>13,17</sup> (red). The estimated susceptible fraction of the population is much lower under the TSIRS model, and grows with time. By contrast, under the models introduced here, rapid waning of immunity allows the susceptible fraction to return seasonally to high levels. The predicted high prevalence of inapparent infection implies that the susceptible pool is rapidly depleted during an outbreak. **b**, Simulated monthly deaths from the TSIRS model. Despite the low susceptible fraction and that model's non-mass-action transmission assumption, the relatively low prevalence of inapparent infection leads to overly explosive outbreaks and overly deep crashes. **c**, Simulated monthly deaths from the SIRS model with seasonal reservoir. **d**, Cholera death data from Dacca district. Parameter estimates for the seasonal-reservoir SIRS model are reported in the Supplementary Information; parameters for the TSIRS model are from K. Koelle (personal communication). The semi-parametric regression used to fit the TSIRS model is initialized using the first 11 yr of data; no simulated data are therefore generated over that interval.

1. Kaper, J. B., Morris, J. G. & Levine, M. M. Cholera. *Clin. Microbiol. Rev.* **8**, 48–86 (1995).
2. Feachem, R. G. Environmental aspects of cholera epidemiology. III. Transmission and control. *Trop. Dis. Bull.* **79**, 1–47 (1982).
3. Van De Linde, P. A. M. & Forbes, G. I. Observations on the spread of cholera in Hong Kong, 1961–63. *Bull. World Health Organ.* **32**, 515–530 (1965).
4. McCormack, W. M., Islam, M. S., Fahimuddin, M. & Mosley, W. H. A community study of inapparent cholera infections. *Am. J. Epidemiol.* **89**, 658–664 (1969).
5. Glass, R. I. & Black, R. E. in *Cholera* (eds Barua, D. & Greenough, W. B. III) 129–154 (Plenum Medical Book Co., New York, 1992).
6. Islam, M. S., Drasar, B. S. & Sack, R. B. The aquatic environment as a reservoir of *Vibrio cholerae*: a review. *J. Diarrhoeal Dis. Res.* **11**, 197–206 (1993).
7. Ionides, E. L., Bretó, C. & King, A. A. Inference for nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **103**, 18438–18443 (2006).
8. Sack, D. A., Sack, R. B., Nair, G. B. & Siddique, A. K. Cholera. *Lancet* **363**, 223–233 (2004).
9. Sanitary Commissioner for Bengal Reports and Bengal Public Health Reports. (Bengal Secretariat Press, Calcutta and Bengal Government Press, Alipore, 1891–, 1942).
10. Glass, R. I., Lee, J. V., Huq, M. I., Hossain, K. M. & Khan, M. R. Phage types of *Vibrio cholerae* O1 biotype El Tor isolated from patients and family contacts in Bangladesh: epidemiologic implications. *J. Infect. Dis.* **148**, 998–1004 (1983).
11. Sack, R. B. *et al.* A 4-year study of the epidemiology of *Vibrio cholerae* in four rural areas of Bangladesh. *J. Infect. Dis.* **187**, 96–101 (2003).
12. Huq, A. *et al.* Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Appl. Environ. Microbiol.* **71**, 4645–4654 (2005).
13. Koelle, K. & Pascual, M. Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera. *Am. Nat.* **163**, 901–913 (2004).
14. Levine, M. M. *et al.* in *Acute Enteric Infections in Children: New Prospects for Treatment and Prevention* (eds Holme, T., Holmgren, J., Merson, M. H. & Mollby, R.) 443–459 (Elsevier/North-Holland Biomedical Press, Amsterdam, 1981).
15. Cash, R. A. *et al.* Response of man to infection with *Vibrio cholerae*. II. Protection from illness afforded by previous disease and vaccine. *J. Infect. Dis.* **130**, 325–333 (1974).
16. Glass, R. I. *et al.* Endemic cholera in rural Bangladesh, 1966–1980. *Am. J. Epidemiol.* **116**, 959–970 (1982).
17. Koelle, K., Rodo, X., Pascual, M., Yunus, M. & Mostafa, G. Refractory periods and climate forcing in cholera dynamics. *Nature* **436**, 696–700 (2005).

18. World Health Organization. *Wkly Epidemiol. Rec.* (<http://www.who.int/wer>) (1950–, 1965).
19. Cash, R. A. *et al.* Response of man to infection with *Vibrio cholerae*. I. Clinical, serologic, and bacteriologic responses to a known inoculum. *J. Infect. Dis.* **129**, 45–52 (1974).
20. Hartley, D. M., Morris, J. G. & Smith, D. L. Hyperinfectivity: a critical element in the ability of *V. cholerae* to cause epidemics? *PLoS Med.* **3**, 63–69 (2006).
21. Siddique, A. K. *et al.* Survival of classic cholera in Bangladesh. *Lancet* **337**, 1125–1127 (1991).
22. Bouma, M. J. & Pascual, M. Seasonal and interannual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologia* **460**, 147–156 (2001).
23. Faruque, S. M. *et al.* Self-limiting nature of seasonal cholera epidemics: Role of host-mediated amplification of phage. *Proc. Natl Acad. Sci. USA* **102**, 6119–6124 (2005).
24. Faruque, S. M. *et al.* Seasonal epidemics of cholera inversely correlate with the prevalence of environmental cholera phages. *Proc. Natl Acad. Sci. USA* **102**, 1702–1707 (2005).
25. Jensen, M. A., Faruque, S. M., Mekalanos, J. J. & Levin, B. R. Modeling the role of bacteriophage in the control of cholera outbreaks. *Proc. Natl Acad. Sci. USA* **103**, 4652–4657 (2006).
26. R Development Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria, 2007).
27. Rodo, X., Pascual, M., Fuchs, G. & Faruque, A. S. G. ENSO and cholera: a nonstationary link related to climate change? *Proc. Natl Acad. Sci. USA* **99**, 12901–12906 (2002).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge discussions with K. Koelle, C. Bretó and A.P. Dobson. This work was funded by the joint National Science Foundation–National Institutes of Health Ecology of Infectious Diseases Program (NSF grants nos 0545276 and 0430120) and the National Oceanic and Atmospheric Administration (Oceans and Health, grant no. NAO40AR460019).

**Author Contributions** A.A.K. formulated and implemented the models, implemented the fitting algorithm, performed the analyses, and drafted and revised the text. E.L.I. provided input on the models and statistical analyses, and drafted much of the Supplementary Information. M.P. provided input on the models and commented on the text. M.J.B. extracted and assembled the data and provided input on the models and on the historical and clinical aspects of cholera.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to A.A.K. ([kingaa@umich.edu](mailto:kingaa@umich.edu)).

## METHODS

**Inference for dynamical systems.** A stochastic dynamical system is generated by a suitable function  $f(x, s, t, \theta, W)$ . Here,  $W$  is a stochastic quantity that is drawn independently each time the function is evaluated and  $\theta$  is a vector of parameters. Thus, for a sequence of times  $t_0 \leq t_1 \leq \dots \leq t_N$ , the recursion

$$X(t_n) = f(X(t_{n-1}), t_{n-1}, t_n, \theta, W),$$

together with the initial value  $X(t_0)$ , specifies a Markov process. The unspecified nature of  $f(\cdot)$  reflects the idea that the evolution of the system could correspond to an essentially arbitrary stochastic simulation algorithm. In the context of this paper,  $X(t)$  is a vector of counts in each disease class and  $f(\cdot)$  corresponds to a (numerical) solution of any of the stochastic differential equation systems in the Supplementary Equations. The data consist of observations  $y_1, \dots, y_n$  at times  $t_1, \dots, t_n$ . We suppose that, conditional on  $\{X(t_k) = x_k\}_{k=1}^n, y_n$  is drawn from a density  $g(y_n | x_n, t_n, \theta)$ .

In the Supplementary Methods, we present a pseudocode description of an iterated filtering methodology (maximization by iterated filtering, MIF) which has recently been shown to enable likelihood-based inference for this framework<sup>7</sup>. This procedure allows inference for rather general partially observed, stochastic, multivariate, continuous-time dynamical systems. The MIF algorithm also accommodates measurement error, non-stationarity, irregular observation times and the mechanistic inclusion of covariates. It has been implemented in the R statistical language<sup>26</sup> as part of the package *pomp*. Note that the algorithmic parameters play a role in the computational efficiency of the method but do not affect the scientific conclusions once likelihood maximization has been confirmed by diagnostic plots<sup>7</sup>. The package and all other codes will be made available by the first author on request.

**Fitting the Bengal data.** Our statistical inference technique is based on simulation of the stochastic dynamical system. To simulate the stochastic dynamical systems, we used the Euler–Maruyama numerical scheme<sup>28</sup> with a timestep of 0.05 months. We fit the parameters of each model to the data from each district using the algorithm given in the Supplementary Methods. We used  $J = 10,000$  particles, a fixed lag  $L = 60$  and  $M = 80$  iterations with a cooling factor,  $a = 0.95$ . Using the same algorithm, we then constructed profile likelihoods against the parameters  $\varepsilon$  and  $c$ . This analysis revealed the existence of weak non-identifiabilities among the parameters  $c, \gamma, m, \varepsilon$  and  $\rho$ . Despite this, certain combinations of parameters, such as  $R_0 = c(\beta_{\text{seas}}(t))/(\gamma + \delta + m)$  and the case fatality  $m/(\gamma + \delta + m)$ , are well-identified. Moreover, although estimates of the durations of short-term immunity vary considerably among districts, it is the case that, for all districts, the duration of immunity is predicted to be far shorter than that derived from severe infections induced experimentally<sup>13–17</sup>. Construction of the profile likelihoods also facilitated the location of the global maximum likelihood estimates, which, after further refinement using the algorithm above with 30,000 particles, are reported in the Supplementary Tables.

**Review of likelihood-based inference.** We give a review of likelihood-based inference in the context of this paper. This review focuses on practical considerations; we refer the reader to Casella & Berger<sup>29</sup> or Rice<sup>30</sup> for a complete introduction, and to Barndorff-Nielsen & Cox<sup>31</sup> for a more advanced treatment. For a discussion of the role of likelihood, profile likelihood ratio tests applied to stochastic models for biological systems, we direct the reader to Hilborn & Mangel<sup>32</sup>. Reasons to use likelihood as a basis for inference include statistical efficiency (that is, making the strongest conclusions available from limited data), objectivity (that is, providing a single criterion applicable to a wide range of models and data) and the availability of useful techniques for comparing rival hypotheses.

For time-series data, written as  $x_{1:T} = x_1, \dots, x_T$ , a stochastic model consists of a density function  $f(x_{1:T} | \theta)$  depending on a vector  $\theta$  of unknown parameters. The likelihood function arises from considering the density as a function of  $\theta$ , treating the data as fixed. Thus the log likelihood is defined to be

$$\log \mathcal{L}(\theta) = \log f(x_{1:T} | \theta),$$

where  $\log$  is the natural logarithm. Rival hypotheses, labelled  $H_0$  and  $H_1$ , concerning the appropriate structure for a stochastic model can be formalized in terms of whether  $\theta$  takes a value in  $\Theta_0$  or  $\Theta_1$  respectively, where  $\Theta_0$  and  $\Theta_1$  are subsets of the set  $\Theta$  of all possible parameter vectors. We refer to  $H_0$  and  $H_1$  as hypotheses or models interchangeably.

**Comparison of nested models.** Suppose that  $H_0$  and  $H_1$  are nested, meaning that  $\Theta_0$  is a subset of  $\Theta_1$ . In practical terms, this is taken to mean that  $\Theta_0$  is described by  $d_0$  freely varying parameters, whereas  $\Theta_1$  is described by these  $d_0$  parameters together with a number  $d_1 - d_0$  of additional parameters. Thus  $\Theta_0$  can be written in terms of a vector of length  $d_0$  and  $\Theta_1$  in terms of a vector of length  $d_1$ . The

maximum likelihood parameter estimate for model  $H_0$  is denoted by  $\hat{\theta}_0$ , and corresponds to the parameter value giving rise to the largest value of  $\log \mathcal{L}(\theta)$  among all values of  $\theta$  in  $\Theta_0$ . A corresponding definition applies to  $\hat{\theta}_1$ . Because  $\Theta_1$  includes  $\Theta_0$ , it follows that  $\log \mathcal{L}(\hat{\theta}_1) \geq \log \mathcal{L}(\hat{\theta}_0)$ . Under standard regularity conditions, twice the improvement in the log likelihood that occurs when moving from the simpler model  $H_0$  to the more complex model  $H_1$  can be compared with a  $\chi^2$  random variable on  $d_1 - d_0$  degrees of freedom. Thus, Supplementary Table 1 presents a  $P$ -value of the form

$$P = \text{Prob} \left[ \chi_{d_1 - d_0}^2 > 2 \left\{ \log \mathcal{L}(\hat{\theta}_1) - \log \mathcal{L}(\hat{\theta}_0) \right\} \right]. \quad (1)$$

The  $P$ -value given by equation (1) is an approximation. More precise  $P$ -values can in principle be obtained by simulation, but this  $\chi^2$  approximation is a widely used technique that has been found to be reliable for many practical purposes<sup>33</sup>. It is both conceptually and operationally simpler to calculate initially the standard  $\chi^2$  approximation in equation (1); if this gives rise to an unequivocal answer, there is little reason to present a simulation study as additional evidence.

Modifications to equation (1) are required when the models under consideration are nested in a non-standard way<sup>34</sup>. This arises in the context of the analysis presented in Supplementary Table 1, owing to the non-standard nesting of the two-path model within the SIRS model. Specifically, the two-path model reduces to the SIRS model when  $c = 1$  at which point  $\rho$  is unidentified. Anisimova *et al.*<sup>35</sup> investigated similar situations and demonstrated that the chi-squared approximation to the likelihood ratio test is conservative in such cases. In the caption of Supplementary Table 1, we report the  $P$ -value associated with such a test on three degrees of freedom:  $\rho, c$  and the initial condition  $Y_0$ .

Non-nested models may still be compared by their likelihood, for example by the use of Akaike's information criterion<sup>36,37</sup>, but formal testing of hypotheses becomes more awkward. Thus, in Supplementary Table 1, we contrast the best-fitting SIRS model with the nested SIRS model constrained to match the duration of immunity found by Koelle & Pascual<sup>13</sup>. Another consideration motivating this comparison is that the nested SIRS model has a similar likelihood to the model of Koelle & Pascual<sup>13</sup>, while having fewer parameters to estimate. Additionally, this approach avoids the issue that the comparison with previously published models might be unfair because the procedure used to fit the models in this paper may be more effective than that used by previous research.

**Profile likelihood.** The profile likelihood is a description of the weight of evidence about the value of a single component of the vector of unknown parameters. Setting  $\theta = (\theta_1, \dots, \theta_d)$ , the log likelihood can be written as  $\log \mathcal{L}(\theta_1, \dots, \theta_d)$  and the profile log likelihood of  $\theta_1$  is

$$\log \mathcal{L}_{\text{profile}}(\theta_1) = \max_{\theta_2, \dots, \theta_d} \log \mathcal{L}(\theta_1, \dots, \theta_d).$$

From equation (1), a 95% confidence region for  $\theta_1$  is given by the range of values of  $\theta_1$  for which

$$2 \left( \log \mathcal{L}(\hat{\theta}) - \log \mathcal{L}_{\text{profile}}(\theta_1) \right) < c,$$

where  $c$  is defined by  $\text{Prob}[\chi_1^2 < c] = 0.95$ . This is how the confidence interval in Fig. 2 was constructed.

28. Kloeden, P. E. & Platen, E. *Numerical Solution of Stochastic Differential Equations* 3rd edn (Springer, New York, 1999).
29. Casella, G. & Berger, R. L. *Statistical Inference* (Wadsworth, Pacific Grove, 1990).
30. Rice, J. A. *Mathematical Statistics and Data Analysis* (Wadsworth, Belmont, California, 1988).
31. Barndorff-Nielsen, O. E. & Cox, D. R. *Inference and Asymptotics* (Chapman and Hall, London, 1994).
32. Hilborn, R. & Mangel, M. *The Ecological Detective* (Princeton University Press, Princeton, New Jersey, 1997).
33. McCullagh, P. & Nelder, J. A. *Generalized Linear Models* 2nd edn (Chapman and Hall, London, 1989).
34. Self, S. G. & Liang, K.-Y. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* **82**, 605–610 (1987).
35. Anisimova, M., Bielawski, J. P. & Yang, Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**, 1585–1592 (2001).
36. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
37. Burnham, K. P. & Anderson, D. R. *Model Selection and Inference: A Practical Information-Theoretic Approach* 2nd edn (Springer, New York, 2002).

# Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice

James F. A. Poulet<sup>1</sup> & Carl C. H. Petersen<sup>1</sup>

Internal brain states form key determinants for sensory perception, sensorimotor coordination and learning<sup>1,2</sup>. A prominent reflection of different brain states in the mammalian central nervous system is the presence of distinct patterns of cortical synchrony, as revealed by extracellular recordings of the electroencephalogram, local field potential and action potentials. Such temporal correlations of cortical activity are thought to be fundamental mechanisms of neuronal computation<sup>3–11</sup>. However, it is unknown how cortical synchrony is reflected in the intracellular membrane potential ( $V_m$ ) dynamics of behaving animals. Here we show, using dual whole-cell recordings from layer 2/3 primary somatosensory barrel cortex in behaving mice, that the  $V_m$  of nearby neurons is highly correlated during quiet wakefulness. However, when the mouse is whisking, an internally generated state change reduces the  $V_m$  correlation, resulting in a desynchronized local field potential and electroencephalogram. Action potential activity was sparse during both quiet wakefulness and active whisking. Single action potentials were driven by a large, brief and specific excitatory input that was not present in the  $V_m$  of neighbouring cells. Action potential initiation occurs with a higher signal-to-noise ratio during active whisking than during quiet periods. Therefore, we show that an internal brain state dynamically regulates cortical membrane potential synchrony during behaviour and defines different modes of cortical processing.

To explore their environment actively, mice rapidly move their whiskers back and forth at about 8–12 Hz in a stereotypical behaviour termed whisking<sup>2,12</sup>. Cortical sensory processing differs markedly between such active whisking periods and quiet wakefulness, when the whiskers are not moving<sup>13–15</sup>. The flow of sensory information from receptors in the whisker follicle, via the infraorbital nerve (ION), brainstem and thalamus to the primary somatosensory barrel cortex (Fig. 1a), is therefore strongly regulated by behavioural state. Here we examine the membrane potential dynamics of layer 2/3 neurons in the barrel cortex of behaving mice and use high-speed (500 Hz) filming of whisker movement as a measure of the behavioural state (Fig. 1b).

Neurons in layer 2/3 are of particular interest because their activity can drive behaviour<sup>16</sup> and because they are readily accessible for optical imaging in behaving mice<sup>15,17</sup>. Our recordings revealed prominent changes in cortical dynamics<sup>13</sup>, termed a change in 'brain state', as the quiet, awake mouse begins to whisk (Fig. 1c–h). Slow (1–5 Hz), large-amplitude membrane potential ( $V_m$ ) fluctuations characterize the quiet brain state and are prominent during quiet wakefulness. These slow oscillations are significantly reduced in amplitude during whisking, defining an active brain state (Fig. 1c).

The primary sensory neurons of the IONs, which signal whisker movement to the brain, are silent during quiet wakefulness but fire action potentials (APs) during whisking<sup>18,19</sup>. Therefore, the change in

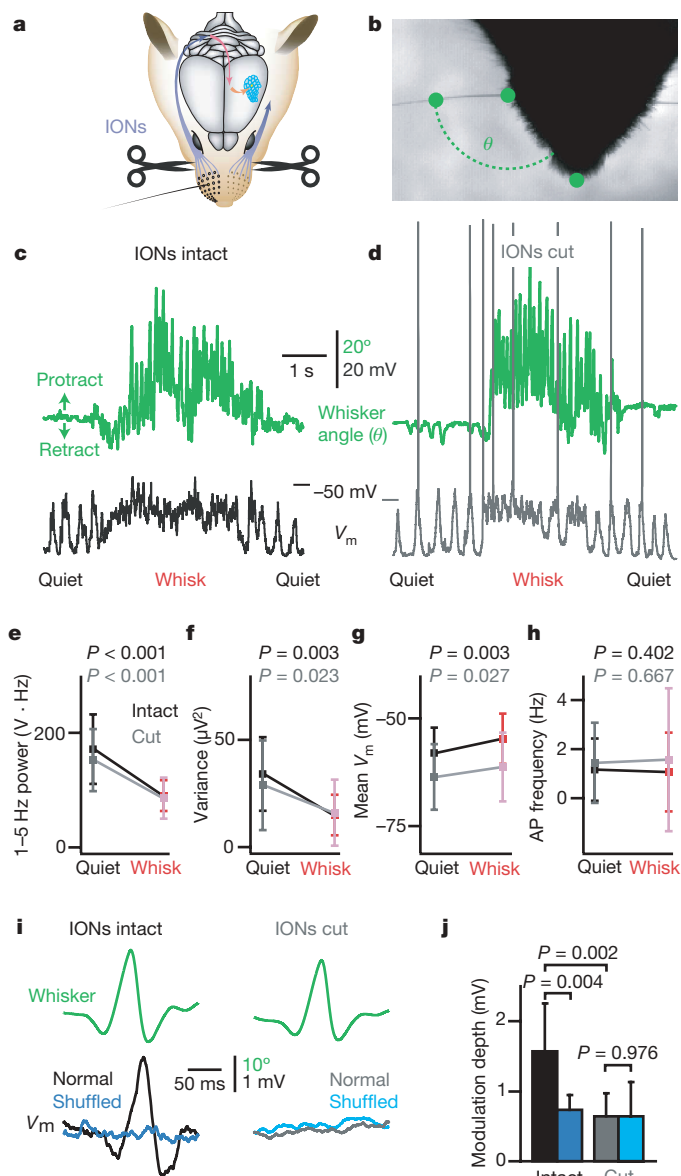
brain state could result from sensory feedback generated during whisker movements, or, alternatively, it could be centrally generated within the central nervous system<sup>20,21</sup>. To distinguish between these two sources, we made bilateral sections of the IONs (Fig. 1a, d). Whisking was not impaired by cutting the IONs, because the facial nerves, containing the motor neuron axons, were left intact. During whisking, mice with intact ( $n = 10$ ) and cut ( $n = 10$ ) sensory nerves showed a strong and significant decrease in the power of low frequency (1–5 Hz) membrane potential oscillations (Fig. 1e) and in the membrane potential variance (Fig. 1f). In both groups of mice, layer 2/3 cortical neurons depolarized slightly but significantly during whisking (Fig. 1g) without overall changes in firing rates (Fig. 1h). None of these parameters was significantly different when comparing preparations in which the nerve was intact to those in which the nerve was cut (Fig. 1e–h). Therefore, the most prominent changes in brain state comparing quiet wakefulness to whisking periods do not depend on sensory input from the whiskers. However, detailed examination of the membrane potential dynamics with respect to the millisecond-by-millisecond trajectory of whisker position revealed differences between mice with intact and cut IONs (Supplementary Fig. 1). Averaging of consecutive whisking cycles aligned to the peak of protraction showed that the membrane potential of layer 2/3 pyramidal neurons in mice with intact sensory nerves exhibited a fluctuation highly correlated with whisker position<sup>13</sup> (Fig. 1i, j). In mice with cut IONs, this fast whisker-position-dependent membrane potential fluctuation was not observed (Fig. 1i, j), in agreement with a previous extracellular study<sup>20</sup>. These data provide functional evidence, alongside the post-mortem anatomical verification, that the IONs were indeed cut. Despite cutting the IONs, the characteristic change in brain state persisted during whisking. We therefore conclude that it must be generated internally, within the central nervous system, and defines an active cortical processing mode.

The key differences between the quiet and the active brain states were the changes in the slow subthreshold membrane potential dynamics. During anaesthesia, slow membrane potential oscillations are highly synchronized<sup>22–24</sup>. We therefore next explored the relationship in behaving mice between single neuron membrane potential dynamics and local network activity measured simultaneously through electroencephalogram (EEG) and the local field potential (LFP; Fig. 2a and Supplementary Figs 2 and 3). The slow large-amplitude fluctuations in  $V_m$  during quiet wakefulness correlated closely with both LFP and EEG measurements (Fig. 2b–d). However, EEG, LFP and  $V_m$  fluctuations decorrelated during active whisking periods (Fig. 2b–d). These state-dependent changes in oscillatory electrical brain activity may be analogous to those first observed in recordings of the human EEG, where subjects relaxing with their eyes closed had slow large-amplitude EEG waves, which disappeared on eye opening<sup>25</sup>. Active processing modes in the neocortex, both in mouse and human, therefore seem to be associated with desynchronized electrical

<sup>1</sup>Laboratory of Sensory Processing, Brain Mind Institute, Faculty of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

activity, whereas large-amplitude slow synchronous oscillations are prominent during rest (7–12 Hz for human and 1–5 Hz for mouse).

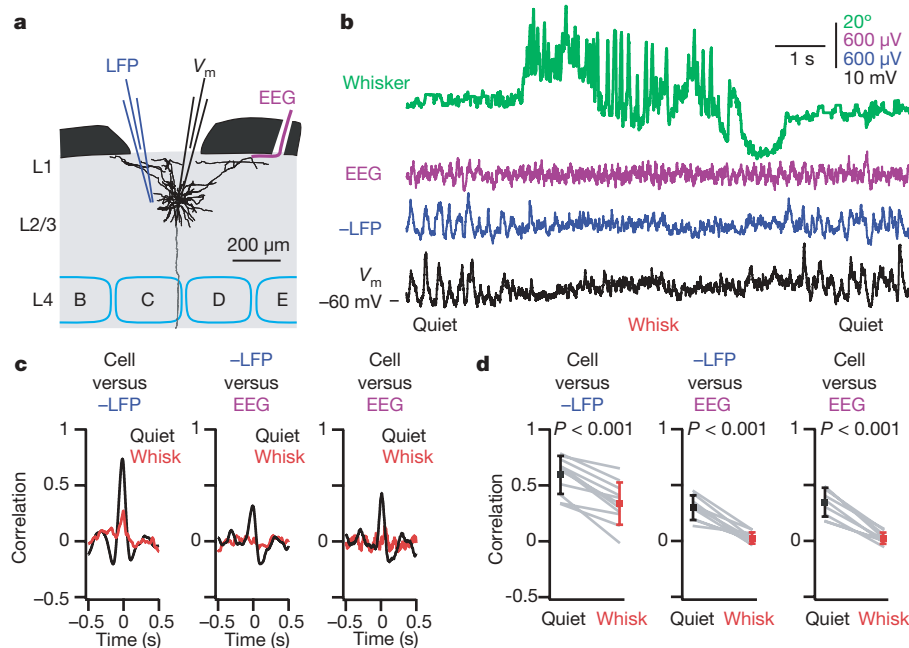
The simplest explanation for the close correlation between LFP and  $V_m$  during quiet wakefulness would be that nearby neurons have highly synchronous slow membrane potential fluctuations. To test this directly, we made simultaneous dual whole-cell recordings from nearby barrel cortex neurons in layer 2/3 (Fig. 3). In an example dual recording of identified layer 2/3 pyramidal neurons (Fig. 3a), membrane potential changes were remarkably similar in the two cells



**Figure 1 | Internal brain state determines membrane potential dynamics.** **a**, Schematic of whisker sensory pathway; scissors indicate position of the bilateral ION cut. **b**, Single image from high-speed (500 Hz) whisker filming showing measurement of whisker angle. **c–h**, Membrane potential ( $V_m$ ) dynamics undergo a change in state during whisking (red) as compared to quiet periods (black), irrespective of whether the IONs were intact (black) or cut (grey). Shown are individual traces from example experiments (**c**, **d**) and quantification across recordings (**e–h**; mean  $\pm$  s.d.;  $n = 10$  intact;  $n = 10$  cut). **i**, Example experiments show an average  $V_m$  modulation phase-locked to whisker movements in a mouse with intact IONs, but not in a mouse with cut IONs. **j**, Averaged across recordings, mice with intact IONs have larger  $V_m$  modulations phase-locked to whisker movement than mice with cut IONs. Mice with intact IONs are different from shuffled data (blue), but mice with cut IONs are not different from shuffled data (mean  $\pm$  s.d.;  $n = 10$  intact;  $n = 10$  cut).

during quiet periods. This is obvious through visual inspection of the overlaid membrane potential traces (Fig. 3b) and was quantified across five different periods in this example experiment by the high cross correlation value of 0.79 (Fig. 3c). Whisking induced a robust and significant decrease in membrane potential correlation in this example experiment to a value of 0.43 (quantified across five periods adjacent to those analysed during quiet wakefulness). A similar pattern of dynamically regulated membrane potential synchrony is shown in a second example experiment (Fig. 3d–f and Supplementary Movie 1; a further example experiment is shown in Supplementary Fig. 4). High correlation values of membrane potential dynamics during quiet periods (mean zero-time-difference cross correlation value  $\pm$  s.d.  $0.72 \pm 0.11$ ,  $n = 19$  mice) and a reduced value during active whisking periods (mean zero-time-difference cross correlation value  $0.33 \pm 0.17$ ,  $n = 19$  mice) were observed in every pair of recorded neurons ( $n = 19$  dual recordings, with a mean of 3.8 whisking episodes analysed per dual recording). The decorrelation induced by the active brain state during whisking was highly significant ( $P < 0.001$ ; Fig. 3g–i); however, the  $V_m$  correlation remains significantly above chance level (cross correlation value of shuffled traces  $0.05 \pm 0.11$ ,  $n = 19$ ,  $P < 0.001$ ). The remaining correlated membrane potential dynamics during whisking might in part be driven by sensory input phase-locked to the whisker cycle (Fig. 1i, j and Supplementary Fig. 5). Fast Fourier transform (FFT) analysis reveals that the most notable changes in the power of membrane potential fluctuations between quiet and active periods occur in the low frequency range (Fig. 3j). Furthermore, spectral coherence analysis indicates that low-frequency membrane potential oscillations also underlie the major state-dependent changes in the amplitude of correlation (Fig. 3k).

These measures of correlated membrane potential fluctuations are entirely dominated by subthreshold activity because APs in layer 2/3 neurons of mouse barrel cortex occur at low rates (mean  $\pm$  s.d.,  $0.61 \pm 0.99$  Hz; median of 0.13 Hz;  $n = 48$ ) both during quiet wakefulness and active whisking periods<sup>13</sup> (Fig. 1h and Supplementary Fig. 6). In our dual whole-cell recordings, we therefore analysed the membrane potential dynamics immediately surrounding individual APs (separated from other APs by at least 100 ms to avoid bursts of APs; Fig. 4 and Supplementary Fig. 6). Whereas membrane potential fluctuations in nearby neurons are highly significantly correlated in both quiet wakefulness and whisking, little or no synchronous AP activity was observed in nearby neurons (Supplementary Fig. 7). The independence of AP firing in nearby neurons provides specificity for information processing. Both during quiet (Fig. 4a, b) and active whisking (Fig. 4c, d) periods, APs were evoked at the same threshold (quiet  $-39.5 \pm 3.3$  mV, whisk  $-39.0 \pm 3.3$  mV,  $n = 9$ ,  $P = 0.20$ ) by a large, brief pre-spike depolarization. Synaptic integration in mouse layer 2/3 pyramidal neurons probably occurs over  $\sim 20$  ms timescale, because this is close to the typical excitatory postsynaptic potential half-width and the membrane time-constant. In the 20 ms preceding AP initiation, the  $V_m$  of the spiking cell depolarized by  $9.0 \pm 2.3$  mV (quiet  $9.7 \pm 2.6$  mV, whisk  $8.3 \pm 2.0$  mV,  $n = 7$ ,  $P = 0.31$ ; Fig. 4e). However, the  $V_m$  of the nearby simultaneously recorded neuron depolarized only by  $1.4 \pm 1.2$  mV (quiet  $1.7 \pm 1.5$  mV, whisk  $1.0 \pm 0.8$  mV,  $n = 7$ ,  $P = 0.28$ ; Fig. 4e). APs are therefore driven by specific synaptic input to a subpopulation of neurons. Although sensory input during whisking (Fig. 1i, j) might contribute to driving these APs, we found a similar value for pre-spike depolarization in mice with cut sensory nerves (quiet  $9.5 \pm 1.7$  mV, whisk  $8.2 \pm 1.8$  mV,  $n = 4$ ). The pre-spike depolarization is large and brief, allowing APs to occur with high temporal precision, irrespective of whether a spike occurs during quiet wakefulness, active whisking or active whisking with cut sensory nerves. It is likely that APs are driven by large pre-spike depolarization, in part, because threshold is lower when approached from a rapidly depolarizing trajectory<sup>26</sup> (Supplementary Fig. 8).

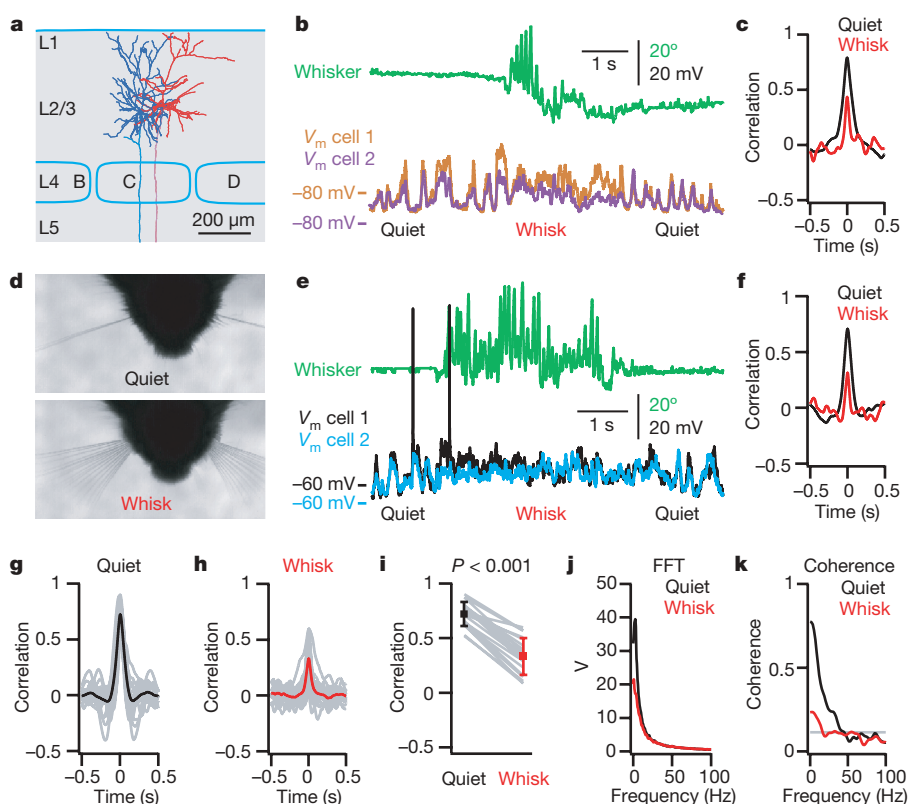


**Figure 2 | EEG, LFP and  $V_m$  are highly correlated during quiet wakefulness but not during active whisking.** **a**, Reconstruction of a layer (L)2/3 pyramidal neuron and schematic showing the recording configuration. **b**, Simultaneous EEG, LFP and whole-cell recordings of the reconstructed neuron oscillate with large amplitude and low frequency during quiet wakefulness and synchronously undergo a state change during whisking. LFP is inverted (up is negative, labelled -LFP). **c**, Correlation analyses across the three recording methods for the example experiment reveals that neuronal activity is more correlated during quiet periods than during active whisking. **d**, Population analysis of the cross correlation across all measures reveals a significant reduction from quiet to whisking. Grey lines indicate different experiments and red/black squares with error bars indicate mean  $\pm$  s.d.;  $n = 11$  for cell versus -LFP;  $n = 9$  for cell versus EEG and for -LFP versus EEG.

We defined a signal-to-noise ratio for AP initiation in layer 2/3 cortical neurons as the ratio of the pre-spike depolarization to the mean variance of the membrane potential fluctuations. This measure shows that APs are driven with significantly higher signal-to-noise ratios during whisking ( $0.54 \pm 0.2$ ;  $n = 7$ ) compared to quiet wakefulness ( $0.24 \pm 0.11$ ;  $n = 7$ ;  $P = 0.008$ ; Fig. 4f) when the pre-spike depolarization is measured 20 ms before threshold. This signal-to-noise ratio measure remains significantly higher during whisking compared to quiet wakefulness when pre-spike depolarizations are measured 5 ms, 10 ms and 50 ms before threshold ( $P < 0.01$  in each case).

A large, brief and specific input therefore drives APs in layer 2/3 barrel cortex neurons of awake mice. Such membrane potential dynamics appear well-suited for the reliable coding of information in networks with sparse AP activity<sup>13,16,27–30</sup>. Future experiments must determine whether similar principles govern APs in other cortical layers and other cell types.

Correlated neuronal activity was first observed in EEG recordings of the human brain, revealing slow, large-amplitude oscillations during periods of quiet wakefulness<sup>25</sup>. Since then, extracellular electrophysiological studies have found evidence for synchronous activity at frequencies ranging from 0.05 Hz to 500 Hz, correlating closely with



**Figure 3 | Subthreshold membrane potential synchrony in layer 2/3 barrel cortex depends on the internal brain state.** **a**, Anatomical reconstruction of two layer 2/3 pyramidal neurons of the mouse barrel cortex. **b**, Simultaneous dual whole-cell recording of these nearby pyramidal neurons during quiet and whisking behaviour. **c**, Cross correlation of  $V_m$  in these two neurons shows high correlation during quiet wakefulness and a reduced correlation during whisking. **d**, Superimposed photos showing whisker movement during whisking. **e, f**, A second example of a pair of layer 2/3 cortical neurons during quiet and whisking highlights the reduction in peak correlation from quiet to whisking. **g, h**, Superimposed cross correlations from 19 pairs of simultaneously recorded neurons (grey) during quiet (g) and whisking (h) behaviour, with the population average highlighted in black and red, respectively. **i**, The cross-correlation value is significantly lower during whisking compared to quiet wakefulness (mean  $\pm$  s.d.;  $n = 19$ ). **j**, Average FFT of  $V_m$  from all neurons shows a reduction in low frequency power during whisking. **k**, Coherence analysis across all dual whole-cell recordings during quiet and whisking indicates low frequency synchronous  $V_m$  fluctuations. The grey line indicates 95% confidence level.

changes in behavioural state, attention, sensory processing and motor planning<sup>1–11</sup>. Investigation of the underlying subthreshold events generating these patterns of cortical activity requires simultaneous dual intracellular recordings of membrane potential. Previous studies have shown that nearby cortical neurons exhibit highly synchronous membrane potential fluctuations under anaesthesia<sup>22–24</sup>; however, dual intracellular recordings have never previously been made in behaving mammals. Here we have shown that a change in brain state dynamically and profoundly regulates cortical membrane potential synchrony during behaviour. The change in brain state is regulated by an internally generated signal, because it persists after cutting the sensory nerves. In future studies, it will be of great interest to define the underlying central mechanisms generating the brain state changes in somatosensory cortex during whisking, and to determine whether they relate to direct effects of neuromodulators in the neocortex or to changes in corticocortical or thalamocortical input.

Our data show that neighbouring layer 2/3 neurons become more independent during the active state of the mouse barrel cortex. Active cortical brain states may therefore serve to augment the total cortical

information processing capacity through decorrelation of membrane potential synchrony while increasing signal-to-noise ratios for AP initiation.

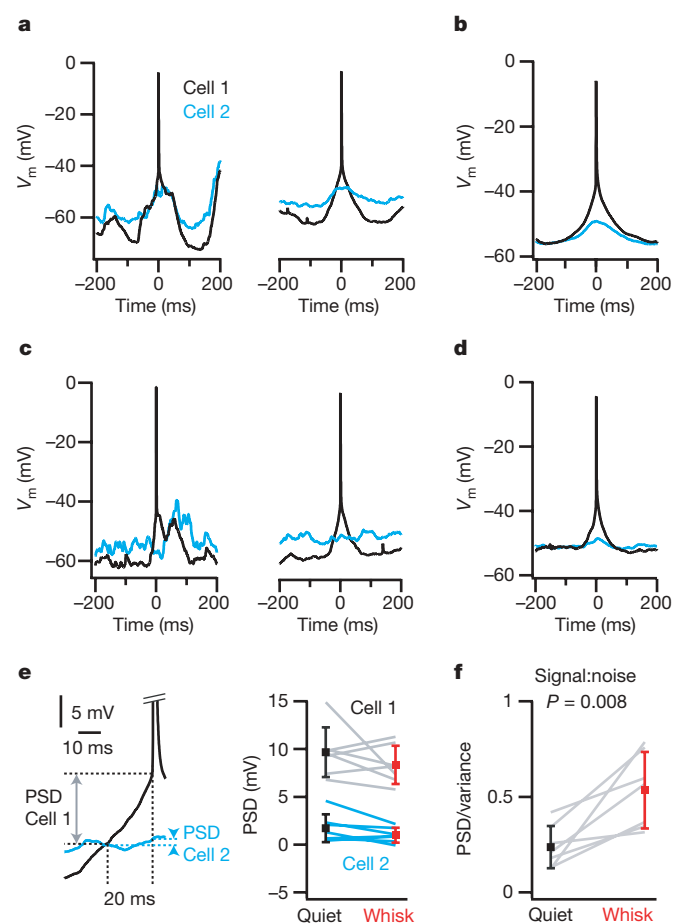
## METHODS SUMMARY

Male 4–6-week-old C57BL/6J mice were implanted with a light-weight metal head holder and a recording chamber. They were then habituated to head restraint in the recording setup over a period of several days. All whiskers, except the C2 whiskers, were trimmed and the location of the left C2 barrel column was functionally located through intrinsic optical imaging, which was used to target the recordings through a small craniotomy<sup>13</sup>. Bilateral sectioning of the infra-orbital nerves was completed 2–4 h before neuronal recording and was confirmed post-mortem. Whole-cell pipettes had resistances of 5–7 M $\Omega$  and were filled with a solution containing (in mM): 135 potassium gluconate, 4 KCl, 10 HEPES, 10 phosphocreatine, 4 MgATP and 0.3 Na<sub>3</sub>GTP (adjusted to pH 7.3 with KOH), as well as 2 mg ml<sup>-1</sup> biocytin to allow anatomical reconstruction of cell morphology and location. Only neurons recorded in layer 2/3 were included in this study. LFP recordings in the immediate vicinity of the whole-cell recordings were made with 5–7 M $\Omega$  glass electrodes filled with Ringer's solution. EEG recordings were obtained using a metal wire recording electrode on the surface of the barrel cortex and a reference electrode on the cerebellum. We used a high-speed (500 Hz) camera to film whisker movements under infrared illumination and custom-written software to measure whisker angle. Data are presented as mean  $\pm$  standard deviation. Statistical significance was assessed using Student's *t*-tests (paired or unpaired assuming unequal variances).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 27 November 2007; accepted 6 June 2008.

Published online 16 July 2008.



**Figure 4 | Action potentials result from large, brief and specific input during both quiet wakefulness and whisking.** **a, c**, Simultaneous membrane potential recording from two layer 2/3 neurons during quiet wakefulness (**a**) and whisking (**c**), focusing on single APs in one of the neurons, which we arbitrarily label cell 1 (black traces). The  $V_m$  in the nearby neuron, which we label cell 2 (blue traces), remains subthreshold, indicating a lack of AP synchrony. A single example (left) is accompanied by the spike-triggered average for this cell pair (right). **b, d**, The spike-triggered grand average  $V_m$  across dual recordings during quiet wakefulness (**b**) and whisking (**d**) ( $n = 7$ ). **e**, The pre-spike depolarization (PSD) quantifies the change in  $V_m$  of both cells during the 20 ms preceding AP initiation in cell 1 (mean  $\pm$  s.d.;  $n = 7$ ). **f**, The ratio of the depolarization driving APs (PSD) relative to the variance of ongoing  $V_m$  fluctuations (or signal-to-noise ratio for evoking a spike) is higher during whisking than during quiet wakefulness (mean  $\pm$  s.d.;  $n = 7$ ).

- Gilbert, C. D. & Sigman, M. Brain states: top-down influences in sensory processing. *Neuron* **54**, 677–696 (2007).
- Petersen, C. C. H. The functional organization of the barrel cortex. *Neuron* **56**, 339–355 (2007).
- Buzsaki, G. & Draguhn, A. Neuronal oscillations in cortical networks. *Science* **304**, 1926–1929 (2004).
- Engel, A. K., Fries, P. & Singer, W. Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Rev. Neurosci.* **2**, 704–716 (2001).
- Gray, C. M., König, P., Engel, A. K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* **338**, 334–337 (1989).
- Logothetis, N. K., Kayser, C. & Oeltermann, A. *In vivo* measurement of cortical impedance spectrum in monkeys: implications for signal propagation. *Neuron* **55**, 809–823 (2007).
- Riehle, A., Grun, S., Diesmann, M. & Aertsen, A. Spike synchronization and rate modulation differentially involved in motor cortical function. *Science* **278**, 1950–1953 (1997).
- Salinas, E. & Sejnowski, T. J. Correlated neuronal activity and the flow of neural information. *Nature Rev. Neurosci.* **2**, 539–550 (2001).
- Steinmetz, P. N. et al. Attention modulates synchronized neuronal firing in primate somatosensory cortex. *Nature* **404**, 187–190 (2000).
- Steriade, M., McCormick, D. A. & Sejnowski, T. J. Thalamocortical oscillations in the sleeping and aroused brain. *Science* **262**, 679–685 (1993).
- Timofeev, I., Grenier, F. & Steriade, M. Disfacilitation and active inhibition in the neocortex during the natural sleep-wake cycle: an intracellular study. *Proc. Natl Acad. Sci. USA* **98**, 1924–1929 (2001).
- Kleinfeld, D., Ahissar, E. & Diamond, M. E. Active sensation: insights from the rodent vibrissa sensorimotor system. *Curr. Opin. Neurobiol.* **16**, 435–444 (2006).
- Crochet, S. & Petersen, C. C. H. Correlating whisker behavior with membrane potential in barrel cortex of awake mice. *Nature Neurosci.* **9**, 608–610 (2006).
- Castro-Alamancos, M. A. Absence of rapid sensory adaptation in neocortex during information processing states. *Neuron* **41**, 455–464 (2004).
- Ferezou, I. et al. Spatiotemporal dynamics of cortical sensorimotor integration in behaving mice. *Neuron* **56**, 907–923 (2007).
- Huber, D. et al. Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice. *Nature* **451**, 61–64 (2008).
- Dombeck, D. A., Khabbazi, A. N., Collman, F., Adelman, T. L. & Tank, D. W. Imaging large-scale neural activity with cellular resolution in awake, mobile mice. *Neuron* **56**, 43–57 (2007).
- Leiser, S. C. & Moxon, K. A. Responses of trigeminal ganglion neurons during natural whisking behaviors in the awake rat. *Neuron* **53**, 117–133 (2007).
- Szwed, M., Bagdasarian, K. & Ahissar, E. Encoding of vibrissa active touch. *Neuron* **40**, 621–630 (2003).
- Fee, M. S., Mitra, P. P. & Kleinfeld, D. Central versus peripheral determinants of patterned spike activity in rat vibrissa cortex during whisking. *J. Neurophysiol.* **78**, 1144–1149 (1997).

21. Ahrens, K. F. & Kleinfeld, D. Current flow in vibrissa motor cortex can phase-lock with exploratory rhythmic whisking in rat. *J. Neurophysiol.* **92**, 1700–1707 (2004).
22. Lampl, I., Reichova, I. & Ferster, D. Synchronous membrane potential fluctuations in neurons of the cat visual cortex. *Neuron* **22**, 361–374 (1999).
23. Petersen, C. C. H., Hahn, T. T. G., Mehta, M., Grinvald, A. & Sakmann, B. Interaction of sensory responses with spontaneous depolarization in layer 2/3 barrel cortex. *Proc. Natl Acad. Sci. USA* **100**, 13638–13643 (2003).
24. Volgushev, M., Chauvette, S., Mukovski, M. & Timofeev, I. Precise long-range synchronization of activity and silence in neocortical neurons during slow-wave oscillations. *J. Neurosci.* **26**, 5665–5672 (2006).
25. Berger, H. Electroencephalogram in humans. *Arch. Psychiatr. Nervenkr.* **87**, 527–570 (1929).
26. Azouz, R. & Gray, C. M. Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons *in vivo*. *Proc. Natl Acad. Sci. USA* **97**, 8110–8115 (2000).
27. Brecht, M., Schneider, M., Sakmann, B. & Margrie, T. W. Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex. *Nature* **427**, 704–710 (2004).
28. Hahnloser, R. H., Kozhevnikov, A. A. & Fee, M. S. An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* **419**, 65–70 (2002).
29. Houweling, A. R. & Brecht, M. Behavioural report of single neuron stimulation in somatosensory cortex. *Nature* **451**, 65–68 (2008).
30. Lee, A. K., Manns, I. D., Sakmann, B. & Brecht, M. Whole-cell recordings in freely moving rats. *Neuron* **51**, 399–407 (2006).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Hill and D. Kleinfeld for help with coherence analyses. We thank S. Crochet, L. Gentet, I. Ferezou and S. Lefort for discussions and critical reading of the manuscript. This work was funded by a Long Term Fellowship from the Human Frontier Science Program and a grant from the Swiss National Science Foundation.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to C.C.H.P. ([carl.petersen@epfl.ch](mailto:carl.petersen@epfl.ch)).

## METHODS

**Animal preparation and surgery.** All experiments were carried out in accordance with the Swiss Federal Veterinary Office, following previously described methods<sup>13</sup>. Male 4–6-week-old, C57BL/6J mice were implanted with a light-weight metal head holder and a recording chamber under 1.5% isoflurane. All whiskers except C2 were trimmed, and the location of the left C2 barrel column was functionally located through intrinsic optical imaging under 1% isoflurane. The cortical surface was visualized through the intact bone. The C2 whisker was deflected at 10 Hz and the evoked haemodynamic signal was imaged by a Qicam CCD camera (Q-imaging) under 630 nm illumination provided by light-emitting diodes. The images were processed online by custom-written routines running in IgorPro (Wavemetrics). A small craniotomy was made over the functionally identified location of the C2 barrel column to target whole-cell and local field potential recordings.

Bilateral sectioning of the IONs was completed 2–4 h before neuronal recording. A small incision was made in the skin exposing the IONs where they emerge from the infraorbital fissure<sup>31</sup>. Sectioning of the IONs was made through the incision starting with the dorsal branches and finishing ventrally. After sectioning the IONs, deflection of the C2 whisker never evoked a detectable intrinsic optical signal. After every recording session, we confirmed complete sectioning of the entire ION on both the left and the right side in a post-mortem examination. The large size of the IONs allows a straightforward verification of the nerve cut. Whisking behaviour was not overtly different between mice with intact nerves and those with cut nerves, as detailed below.

**Habituation to head restraint.** Mice readily habituate to head restraint. The first head-restrained sessions of each mouse lasted only for a few minutes and this time was gradually increased each day until the mouse would sit calmly for a period of roughly 1 h. After several days of habituation to head restraint, recordings were performed in a single session lasting up to 2 h, after which the mouse was killed to obtain anatomical information.

Mice never slept while in head fixation during the training period or during the recording session, and they remained behaviourally responsive to sensory stimuli at all times throughout the recording session. In a previous study, we demonstrated that the EEG patterns observed during active whisking periods, quiet wakefulness and sleep are quantitatively different (Supplementary Fig. 1 in ref. 13).

**Neuronal recordings.** Whole-cell recordings from awake head-restrained mice were obtained as described previously<sup>13</sup>. Pipettes had resistances of 5–7 M $\Omega$  and were filled with a solution containing (in mM): 135 potassium gluconate, 4 KCl, 10 HEPES, 10 phosphocreatine, 4 MgATP and 0.3 Na<sub>3</sub>GTP (adjusted to pH 7.3 with KOH), as well as 2 mg ml<sup>-1</sup> biocytin. Only neurons recorded in layer 2/3 (at depths <500  $\mu$ m below the pia) were included in this study. Whole-cell electrophysiological measurements were made with a Multiclamp 700A amplifier (Axon Instruments). The  $V_m$  was filtered at 10 kHz and digitized at 20 kHz by ITC-18 (Instrutech Corporation) under the control of IgorPro. No current was injected during recording, and  $V_m$  was not corrected for liquid junction potentials.

LFP recordings were made with 5–7 M $\Omega$  glass electrodes filled with Ringer's solution. The LFP electrodes were carefully lowered into layer 2/3 into the same craniotomy and at the same time as the patch electrodes. Depth measurements were <50  $\mu$ m between the LFP and the patch electrode. The LFP signal was amplified by a Multiclamp 700A amplifier, filtered at 10 kHz and digitized at 20 kHz by ITC-18 under the control of IgorPro. LFP recordings were bandpass filtered off-line in IgorPro between 0.1 Hz and 100 Hz.

EEG recordings using exposed tips of 50- $\mu$ m Teflon-coated tungsten wires were obtained from the surface of the barrel cortex (~500  $\mu$ m from the craniotomy) with the reference electrode on the cerebellum. The EEG signal was amplified by CyberAmp 380 amplifier (Axon Instruments), filtered at 0.1 Hz–10 kHz and digitized at 20 kHz by ITC-18 under the control of IgorPro. EEG recordings were filtered off-line in IgorPro between 0.1 Hz and 100 Hz.

**Recording behaviour.** We used a high-speed (500 Hz) MotionPro camera (Redlake) to film whisker movements under infrared illumination. The field of view was 25  $\times$  25 mm, with a single pixel resolution of 0.1 mm. Custom-written image analysis software running in ImageJ was used to measure whisker angle.

Whisking is driven by motoneurons of the facial nucleus, which are controlled by several mechanisms, including serotonergic<sup>32</sup> and cortical control<sup>15,27</sup>. Cortical input is not essential to initiate whisking, because it is clear that whisking can occur in the absence of cortical activity (for example, after lesions to motor cortex)<sup>33,34</sup>. We do not know which control mechanisms give rise to the whisking observed in our study, but we found that the decorrelation of cortical activity occurred robustly in every whisking episode. Therefore, although there might be differences in brain state during the head-fixation session, this did not

impact on our observations. Most bouts of whisking included in this study were initiated in the absence of any overt external event and we therefore consider them to be spontaneous. In addition, we found that auditory, olfactory and visual stimuli could induce whisking. The change in cortical dynamics and synchrony from quiet wakefulness to whisking was robust and did not depend on how whisking was initiated.

Whisking bouts lasted for a similar duration with IONs intact  $4.3 \pm 1.1$  s ( $n = 10$  mice) or IONs cut  $4.5 \pm 1.1$  s ( $n = 10$  mice). FFT analysis of whisker movement showed a peak at 8.5 Hz both for mice with intact ( $n = 10$  mice) and for those with cut ( $n = 10$  mice) IONs.

**Neuron staining.** Each mouse was deeply anaesthetized with urethane and perfused with 4% paraformaldehyde immediately after the experiment. Subsequently, tangential slices were processed for cytochrome oxidase staining to reveal the barrel map and for biocytin staining (Vectastain) to reveal the morphology of the recorded neuron. Neuronal reconstruction was performed in three dimensions with NeuroLucida software (MicroBrightField).

**Data analysis.** Three independent data sets were gathered for Figs 1, 2 and 3. The data shown in Fig. 1 were recorded from 10 neurons in mice with intact IONs and 10 neurons in mice with cut IONs. Each cell was recorded in a different mouse. The data in Fig. 2 were recorded from a further 11 cells in 9 additional mice. The data in Fig. 3 were recorded from 19 dual whole-cell recordings with each pair of neurons in a different additional mouse.

For Fig. 1e–g, properties of membrane potential dynamics were calculated from 3-s segments of data taken from a visually identified central period of each whisking episode during the recordings and compared to adjacent quiet sections of the recording without whisking. Data segments in the nerve-intact configuration were taken from 41 whisking episodes (range 2–6, mean 4.1 per experiment) from 10 experiments; data segments for analysis in the nerve-cut configuration were taken from 56 whisking episodes (range 3–8, mean 5.6 per experiment) from 10 experiments. The power of low frequency  $V_m$  changes was computed by measuring the area under the FFT between 1 Hz and 5 Hz. For Fig. 1h, spike rates were calculated across all quiet and whisking periods in the recorded sweeps. The recording periods analysed for spiking in the 10 intact preparations were: during quiet, total time 729 s (mean  $\pm$  s.d.  $73 \pm 58$  s; range 17–196 s); and during whisking, total time 376 s (mean  $\pm$  s.d.  $38 \pm 20$  s; range 16–71 s). In the 10 experiments in which the nerve was cut, the recording periods analysed for spiking were: during quiet, total time 922 s (mean  $\pm$  s.d.  $92 \pm 57$  s; range 15–170 s); and during whisking, total time 515 s (mean  $\pm$  s.d.  $51 \pm 26$  s; range 15–98 s). For Fig. 1i, every recorded whisking cycle during an experiment was aligned at the peak of protraction and averaged, revealing both the mean whisker movement during a whisking cycle and the mean change in membrane potential. For the example experiment with intact IONs a total of 329 whisking cycles were averaged, and for the example experiment with cut IONs a total of 457 whisking cycles were averaged. For Fig. 1j, modulation depth is defined as the difference in amplitude between the most positive and the most negative peak in the average membrane potential within  $\pm 100$  ms from the peak of protraction. All recorded whisking cycles were included in the average. For the 10 nerve intact preparations we analysed a total of 1,802 whisking cycles (mean  $\pm$  s.d.  $180 \pm 96$ ; range 102–329). For the 10 nerve cut preparations we analysed a total of 2,961 whisking cycles (mean  $\pm$  s.d.  $296 \pm 133$ ; range 132–482).

The mean value was subtracted and the amplitude of fluctuations was normalized to the standard deviation for each individual data segment before computing the cross correlation<sup>22</sup> (Figs 2c, d and 3c, f–i). The cross-correlation value is the mean from  $-2$  ms to  $+2$  ms. For Fig. 2c, d, cross correlations were computed across the central 3-s period of visually identified whisking episodes. In total, 40 whisking episodes (range 1–8, mean 3.6) were analysed in Fig. 2d, with 11 experiments combining whole-cell and local field potential recordings; in 9 of these experiments we additionally measured the EEG. For Fig. 3c, f–i, cross correlations were computed across the central 3-s period of visually identified whisking episodes with a total of 72 episodes (range 2–6, mean 3.8 whisking episodes per dual recording, total 216 s of recording). Cross correlations of quiet periods were computed from the same number of 3-s segments, with each analysed segment being adjacent to the whisking episode. For Fig. 3j, k, the open source Chronux software (<http://chronux.org/>) was used under Matlab (The Mathworks) to generate the FFT and calculate coherence across the 19 dual recordings during quiet and whisking.

For Fig. 4a, d, the average membrane potential of the spiking and nearby cell was calculated from all spikes isolated from another spike by  $>100$  ms. The grand average represents the average from 7 cells. For Fig. 4e, the pre-spike depolarization was calculated as the change in  $V_m$  during the 20 ms before the spiking cell reached threshold for action potential initiation. For Fig. 4f, the signal-to-noise ratio was calculated by dividing the pre-spike depolarization leading to action potential initiation by the variance during quiet and whisking periods observed in the same recording.

All data are presented as mean  $\pm$  standard deviation. Student's *t*-tests (two tailed, paired or unpaired assuming unequal variances) were performed in Excel (Microsoft) to evaluate statistical significance. Unless otherwise stated, the number of cells analysed is denoted by '*n*'.

31. Dörfl, J. The innervation of the mystacial region of the white mouse: a topographical study. *J. Anat.* **142**, 173–184 (1985).
32. Hattox, A., Li, Y. & Keller, A. Serotonin regulates rhythmic whisking. *Neuron* **39**, 343–352 (2003).
33. Gao, P., Hattox, A. M., Jones, L. M., Keller, A. & Zeigler, H. P. Whisker motor cortex ablation and whisker movement patterns. *Somatosens. Mot. Res.* **20**, 191–198 (2003).
34. Welker, W. I. Analysis of sniffing of the albino rat. *Behaviour* **22**, 223–244 (1964).

## LETTERS

## On the spontaneous emergence of cell polarity

Steven J. Altschuler<sup>1</sup>, Sigurd B. Angenent<sup>2</sup>, Yanqin Wang<sup>1</sup> & Lani F. Wu<sup>1</sup>

Diverse cell polarity networks require positive feedback for locally amplifying distributions of signalling molecules at the plasma membrane<sup>1</sup>. Additional mechanisms, such as directed transport<sup>2</sup> or coupled inhibitors<sup>3,4</sup>, have been proposed to be required for reinforcing a unique axis of polarity. Here we analyse a simple model of positive feedback, with strong analogy to the ‘stepping stone’ model of population genetics<sup>5</sup>, in which a single species of diffusible, membrane-bound signalling molecules can self-recruit from a cytoplasmic pool. We identify an intrinsic stochastic mechanism through which positive feedback alone is sufficient to account for the spontaneous establishment of a single site of polarity. We find that the polarization frequency has an inverse dependence on the number of signalling molecules: the frequency of polarization decreases as the number of molecules becomes large. Experimental observation of polarizing Cdc42 in budding yeast is consistent with this prediction. Our work suggests that positive feedback can work alone or with additional mechanisms to create robust cell polarity.

Many cell types can spontaneously establish and maintain asymmetric distributions of signalling molecules on the plasma membrane<sup>1,6</sup>, even in the apparent absence of pre-existing cues<sup>7,8</sup>. Positive feedback circuits, found at the core of diverse biological networks, enable signalling molecules localized at the plasma membrane to initiate processes that further accelerate localized recruitment. These processes allow signalling molecules, such as Cdc42 in budding yeast<sup>9</sup>, mPar3 and mPar6 in axons of hippocampal neurons<sup>10</sup>, and Rac in canine kidney cells<sup>11</sup> and human chemotaxing neutrophils<sup>12</sup>, to be concentrated within a defined region of the plasma membrane.

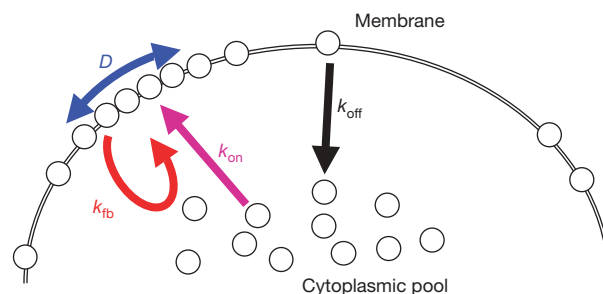
Spontaneous cortical polarization can occur through activation of positive feedback circuits that do or do not depend on mechanisms of directed transport<sup>8,13,14</sup>. Transport-dependent circuits make use of cytoskeletal structure to reinforce spatial asymmetries by directing signalling molecules to specified locations on the plasma membrane<sup>2,15,16</sup>. However, in the absence of directed transport, it is unclear how positive feedback circuits counteract lateral diffusion to establish and maintain an asymmetric distribution of signalling molecules<sup>8,14</sup>. Additional mechanisms have been proposed, such as coupled activation-dependent inhibition<sup>3,4,16</sup> or depletion of critical signalling components<sup>3</sup>. Because evidence for these other mechanisms is less generally established, we wondered whether positive feedback alone could generate cell polarization.

The specific details of any given positive feedback circuit can be complex and cell-type specific. For example, signalling molecules can associate to the plasma membrane through docking with transmembrane receptors, binding to phospholipid headgroups, or insertion of prenylated membrane anchors. Additionally, signalling molecules such as GTPases can assume multiple states, which in turn strongly affect properties of their transport. To discover general principles, as with other recent models of polarity<sup>15,17–19</sup>, we abstracted mechanisms common to these circuits and constructed a biologically motivated, yet mathematically tractable, model of positive feedback.

Thus, we formulated a model of positive feedback in which signalling molecules comprising a single species move between non-recruiting, cytoplasmic states and recruiting, plasma-membrane-bound states without mechanisms of directed transport (this model is referred to simply as ‘positive feedback’ in the remainder of the paper). The redistribution of signalling molecules is determined by the rates of four fundamental transport mechanisms (Fig. 1; see Methods): (1) recruitment ( $k_{fb}$ ) of cytoplasmic molecules to the locations of membrane-bound signalling molecules; (2) spontaneous association ( $k_{on}$ ) of cytoplasmic molecules to random locations on the plasma membrane; (3) lateral diffusion ( $D$ ) of molecules along the membrane; and (4) random disassociation ( $k_{off}$ ) of signalling molecules from the membrane. The model parameters  $k_{fb}$ ,  $k_{on}$ ,  $D$  and  $k_{off}$ , and total number of signalling molecules  $N$ , are readily biologically interpretable, and may be estimated from experimental data (Methods).

In an initial analysis of the dynamics of this positive feedback circuit, we ignored the spatial distribution of molecules on the membrane. We found that the fraction of total signalling molecules on the membrane can switch from zero to a final equilibrium  $h_{eq}$  with a half time of  $\sim 1/(h_{eq}k_{fb})$  (Supplementary Information). This equilibrium fraction  $h_{eq}$  can be estimated simply as  $1 - k_{off}/k_{fb}$  in the case when  $k_{on}$  is smaller than  $k_{fb}$  (see Supplementary Information for general formulation), and reflects a balance between the net fluxes of molecules to and from the membrane. Thus, we find that the strength of feedback strongly influences both the switching dynamics and the magnitude of the equilibrium state.

We next considered how signalling molecules are distributed spatially on the plasma membrane. In the case when the number of signalling molecules is large, analysis reveals that all spatial variation eventually vanishes; membrane densities converge exponentially to a spatially homogeneous equilibrium state with membrane fraction  $h_{eq}$



**Figure 1 | A conceptual model of a positive feedback circuit is characterized by five biologically interpretable parameters.** Coloured arrows indicate four main sources of molecular transport (see text: purple, spontaneous membrane association or ‘input’ to positive feedback circuit, with rate  $k_{on}$ ; black, spontaneous membrane disassociation, with rate  $k_{off}$ ; red, positive feedback, with rate  $k_{fb}$ ; blue, lateral diffusion, with rate  $D$ ). Not shown is  $N$ , the total number of signalling molecules.

<sup>1</sup>Green Center for Systems Biology, Department of Pharmacology and Simmons Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.

<sup>2</sup>Mathematics Department, University of Wisconsin, Madison, Wisconsin 53706, USA.

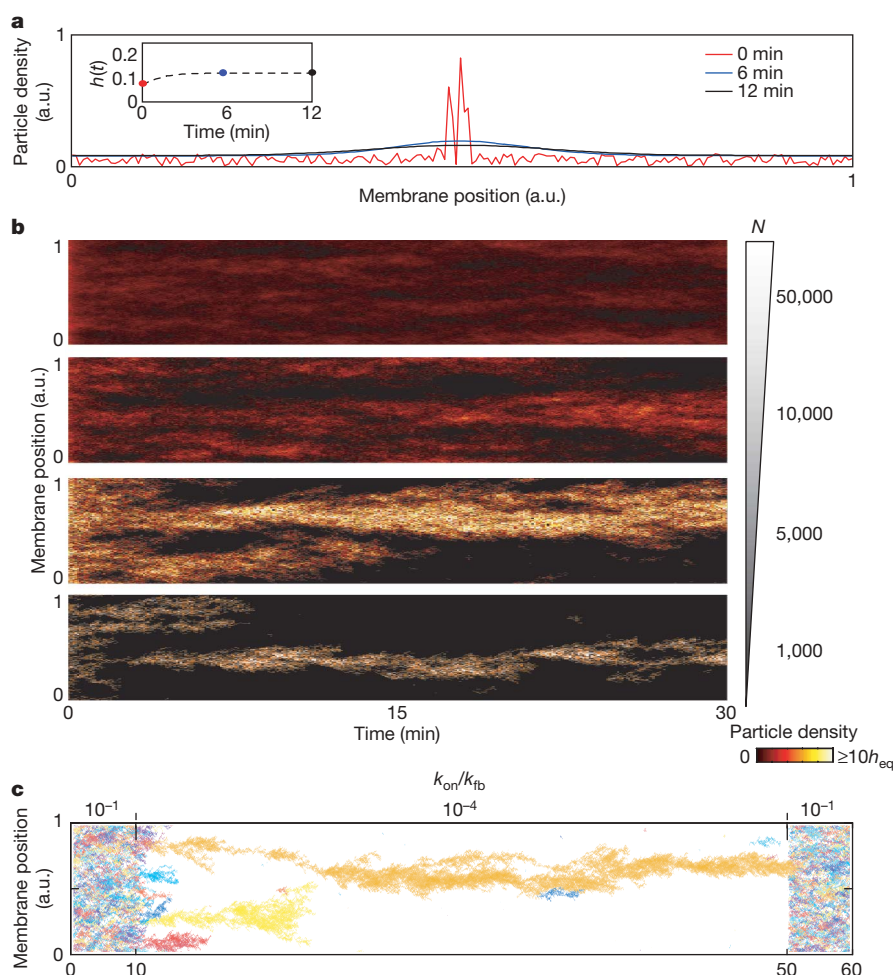
(Methods). Furthermore, perturbations of this state due to spontaneous association events quickly return to equilibrium (Fig. 2a; see Methods). Thus, no polarity can persist, regardless of the feedback strength  $k_{fb}$ . This suggests that for large numbers of molecules, positive feedback requires additional biological mechanisms to overcome diffusion in order to maintain a persistent switch-like distribution in space. This phenomenon has been studied rigorously for other positive feedback systems, including many motivated by theoretical ecology<sup>20</sup>.

In contrast, for small numbers of molecules, unexpected model behaviour emerges as stochastic fluctuations can drive cellular behaviours from stable equilibrium to dynamic, non-stable states<sup>21–24</sup>. As the number of signalling molecules decreases, the expected equilibrium fraction of molecules on the membrane remains  $h_{eq}$ , with fluctuations on the order of  $1/\sqrt{N}$ . However, we observed that with decreased numbers of signalling molecules, distinct polarized regions began to emerge spontaneously (Fig. 2b). Additionally, we observed that this behaviour is strongly affected by the ratio  $k_{on}/k_{fb}$  of spontaneous membrane association to feedback (Fig. 2c and Supplementary Fig. 1). When this ratio is large, frequent association of molecules to random membrane locations results in a spatially homogeneous density (Fig. 2c, first and last 10 min). When this ratio is small, spontaneous association events are rare and local amplification by positive feedback frequently results in the emergence of spatially isolated clusters of signalling molecules (Fig. 2c, middle 40 min). We note that

simulations were performed using  $h_{eq} = 0.1$  (Methods), hence the observed polarization behaviour is not due to depletion of the cytosolic pool (Supplementary Materials)<sup>3</sup>. Thus, these observations predict that decreased numbers of signalling molecules, decreased spontaneous membrane association rate, or increased positive feedback strength can lead to the spontaneous emergence of polarity (Box 1).

Mathematical analysis of the positive feedback system shows that transitions among clustering behaviours are entirely determined by a simple relationship between the ratio  $k_{on}/k_{fb}$  and the number of signalling molecules  $N$ : a single cluster will certainly form for small  $k_{on}/k_{fb} \ll N^{-2}$  (Fig. 3a, region 1), whereas no clusters will form for large  $k_{on}/k_{fb} \gg (N^{-1} \ln N)^{1/2}$  (Fig. 3a, region 2). Numerical simulations indicate that the transition occurs for  $k_{on}/k_{fb} \approx N^{-1}$  (Fig. 3a, region 3). Cluster formation is possible for large numbers of molecules, but the larger the value of  $N$ , the smaller the value of  $k_{on}/k_{fb}$  for which clustering actually occurs. This phase plane portrait reveals that polarization is a robust behaviour that can be tuned by a small number of parameters.

A clear prediction of this theory is that the probability of spontaneous polarization decreases as the number of signalling molecules becomes large (Fig. 3a, follow any horizontal line from left to right). To test whether this trend could be observed experimentally, we made use of an established budding yeast system in which an actin-independent core positive feedback circuit has been shown to be



**Figure 2 | Numerical simulations reveal an inverse dependence between polarization frequency and large numbers of signalling molecules.**

**a**, Positive feedback alone leads to homogeneous steady-state solutions for large numbers of signalling molecules (Methods). Inset: evolution of total membrane fraction  $h(t)$ ; theoretical equilibrium value  $h_{eq} = 0.1$ . a.u., arbitrary units. **b**, Decreasing numbers of signalling molecules lead to

increasing levels of spatial segregation. Shown are kymographs from simulations (y axis, membrane position; x axis, time). Particle density is rescaled to units of equilibrium fraction (colour bar). **c**, Small  $k_{on}/k_{fb}$  can lead to the establishment of a dominant clan (Box 1). Clan colour assignment in the kymograph is random, and is reset for each molecule after membrane disassociation. Model parameters are as in Methods.

**Box 1 | Intuition for particle clustering**

Intuition for the clustering behaviour of positive feedback circuits for small numbers of molecules can be obtained through the simple concept of lineage tracking (Fig. 2c). The basic idea is to track the location and source of recruitment for all signalling molecules on the membrane. Signalling molecules that spontaneously associate with the membrane become the founders of new 'clans'. Molecules recruited to the membrane join the clans of their recruiters. Molecules that leave the membrane to re-enter the cytoplasmic pool lose their clan identity. A clan that has lost all its members becomes extinct.

Initially each clan has population size one, but over time some clans grow whereas others become extinct. Hence the number of clans on the membrane decreases between occurrences of spontaneous on-events. At equilibrium the expected total number of signalling molecules at the membrane remains a (non-zero) constant. If the time between spontaneous on-events is sufficiently long, then the number of clans may decrease so much that all molecules on the membrane belong to a single remaining clan. For small diffusion rates, all molecules on the membrane will be located near their most recent common ancestor, resulting in the formation of a single, high-density cluster. (Our work bears close resemblance to the well studied 'stepping stone' model<sup>5</sup> introduced to study spatially heterogeneous theories of population genetics and dynamics. Clans in the stepping stone model can be more literally interpreted as collections of individuals sharing a common genetic trait by descent; new mutations create new clans, genetic drift increases or decreases clan sizes, and migration disperses clan members among neighbouring colonies.)

How long does it typically take for a single cluster to form at equilibrium? For relatively small membrane association rates we can deduce that the expected number of clans after waiting any time  $t$  is approximately equal to  $h_{eq}N/(k_{off}t)$  (Supplementary Information). Thus, after a time interval of  $t \approx h_{eq}N/k_{off}$ , a single clan is expected if no further spontaneous membrane association occurs. Simulations indicate, however, that a single large clan can form even if a small number of spontaneous associations occur; most new clans rapidly become extinct while the dominant clan is forming. Diffusion eventually spreads apart the members of a dominant clan. However, at any given time, lineage tracking can be restarted with all original clan identity 'erased' and each membrane-bound molecule regarded as a new clan founder. Then, our analysis shows that after a short time a dominant clan will again emerge. Thus, positive feedback alone, operating on a small number of particles, has the capacity to recurrently generate polarity.

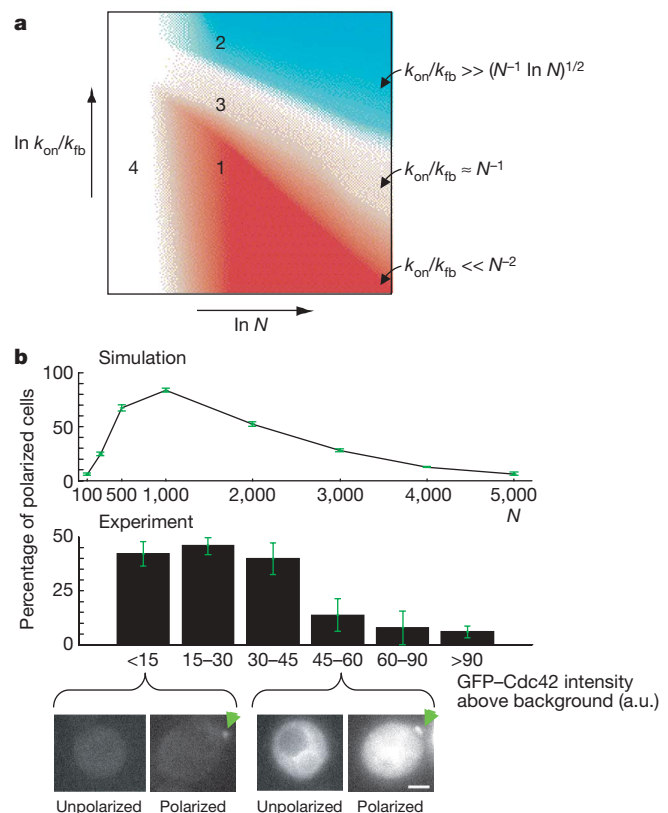
sufficient to establish cue-independent, spontaneous polarization of the GTPase Cdc42 (refs 8, 9, 14). Mechanistic details of this positive feedback circuit are not well understood; however, several lines of evidence indicate its relevance as an experimental model for our current study: (1) self-recruitment of Cdc42 at the plasma membrane (via its interactions with the guanine nucleotide exchange factor (GEF) Cdc24 and the adaptor protein Bem1) does not require directed transport<sup>8,9,14,25,26</sup>; (2) polarization is maintained in a state of dynamic equilibrium, with all components rapidly exchanging between the plasma membrane and a significant cytoplasmic pool<sup>8</sup>; and (3) sites of polarization have been observed occasionally to drift and flicker, in phenomenological agreement with our model simulations<sup>8</sup>.

This positive feedback circuit for Cdc42 can be grossly described by the four fluxes in our model. Model parameter values can be estimated from experimental data (Methods), and reflect the combined effects of presumed unmodelled mechanistic interactions, such as required for self-recruitment (for example, via Bem1 and Cdc24 (refs 8, 9, 14)) and reversible membrane association (for example, via guanine nucleotide disassociation inhibitor (GDI)<sup>8,27–29</sup>). For varying numbers of signalling molecules, probabilities of polarization can be computed using numerical simulation. As expected from the phase plane diagram (Fig. 3a), the probability of spontaneous polarization decreases as the number of signalling molecules becomes large (Fig. 3b, top histogram,  $N > 1,000$ ). We note that as the number of signalling molecules becomes very small the overall polarization

percentage drops sharply (Fig. 3a region 4, and Fig. 3b top histogram,  $N < 500$ ), reflecting more frequent loss of polarity due to spontaneous disassociation of all membrane-bound particles and longer waiting times for re-establishment (Supplementary Fig. 2).

We next assayed polarization of green fluorescent protein (GFP)-tagged Cdc42 in yeast cells treated with the drug latrunculin A<sup>30</sup> to inactivate an alternative actin-dependent feedback circuit<sup>2,8,14,15</sup> (Methods). To quantify the effects of varying  $N$ , cells were binned into different categories on the basis of background-subtracted, total GFP-Cdc42 intensities (Supplementary Fig. 3; see Methods). Within each category we computed the fraction of polarized cells (Fig. 3b). In general agreement with the model prediction, we observed a significant decrease in the frequency of polarization with increasing Cdc42-GFP expression (Fig. 3b, bottom histogram). Notably, this inverse correlation is opposite to what was modelled for transport-dependent positive feedback of constitutively activated Cdc42 in budding yeast. In the case of actomyosin-based feedback, increased concentrations of constitutively activated Cdc42 led to increased frequency of establishing and maintaining distinct sites of cortical polarization<sup>2,15</sup>. Thus, with respect to increasing large numbers of signalling molecules, models suggest opposite trends in the frequency of polarization for positive feedback circuits that do and do not make use of directed feedback.

The intrinsic ability of positive feedback to polarize spontaneously without requiring additional mechanisms is not apparent from



**Figure 3 | Dependence of polarization on model parameters.** **a**, Illustration of phase-plane portrait. Theory reveals robust parameter regimes for cluster formation (Supplementary Information, Summary): 1, polarization; 2, no polarization; 3, transition zone; 4, too few signalling molecules for polarization (Supplementary Fig. 2). **b**, Frequency of observing Cdc42 polarization in yeast for increasing numbers of molecules ( $N$ ). Shown are polarization frequencies estimated from numerical simulations (top panel) and experimental observation of yeast cells expressing GFP-Cdc42 (middle panel) (Methods and Supplementary Fig. 3). Green error bars represent standard error from four independent replicates. Bottom panel: examples of polarized and unpolarized yeast cells. Arrows point to polarized regions. Scale bar, 1.9  $\mu\text{m}$ .

established deterministic models of pattern formation<sup>3,4</sup> (Supplementary Information). Our work suggests that positive feedback can act alone at the core of complex signalling networks with small numbers of molecules to create and maintain a highly localized cortical distribution required for activation of downstream biological processes.

## METHODS SUMMARY

**Mathematical models.** Deterministic and stochastic formulations of the model are given in Methods and Supplementary Information.

**Numerical simulations.** Simulations were performed using parameter values estimated from published experiments<sup>8,15</sup> (Methods):  $D \approx 1.2 \mu\text{m}^2 \text{min}^{-1}$ ;  $N \approx 10^3$ ;  $h_{\text{eq}} \approx 0.1$ ;  $k_{\text{off}} \approx 9 \text{min}^{-1}$ ;  $k_{\text{fb}} \approx 10 \text{min}^{-1}$ ; and  $k_{\text{on}}$  as indicated in each figure. Numerical simulations were carried out on one-dimensional circles.

**Experimental assay.** The presence of GFP-Cdc42 polar caps were assayed in yeast as previously described<sup>8,15</sup> (Supplementary Fig. 3). GFP-Cdc42 intensity scores for each cell region were computed using Image J, and probabilities of cap formation computed using custom Matlab software.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 26 March; accepted 27 May 2008.

- Drubin, D. G. & Nelson, W. J. Origins of cell polarity. *Cell* **84**, 335–344 (1996).
- Wedlich-Soldner, R. *et al.* Spontaneous cell polarization through actomyosin-based delivery of the Cdc42 GTPase. *Science* **299**, 1231–1235 (2003).
- Gierer, A. & Meinhardt, H. A theory of biological pattern formation. *Kybernetik* **12**, 30–39 (1972).
- Turing, A. M. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. B* **237**, 37–72 (1952).
- Kimura, M. & Weiss, G. H. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561–576 (1964).
- Ebersbach, G. & Jacobs-Wagner, C. Exploration into the spatial and temporal mechanisms of bacterial polarity. *Trends Microbiol.* **15**, 101–108 (2007).
- Sohrmann, M. & Peter, M. Polarizing without a c(l)ue. *Trends Cell Biol.* **13**, 526–533 (2003).
- Wedlich-Soldner, R. *et al.* Robust cell polarity is a dynamic state established by coupling transport and GTPase signaling. *J. Cell Biol.* **166**, 889–900 (2004).
- Butty, A. C. *et al.* A positive feedback loop stabilizes the guanine-nucleotide exchange factor Cdc24 at sites of polarization. *EMBO J.* **21**, 1565–1576 (2002).
- Shi, S. H., Jan, L. Y. & Jan, Y. N. Hippocampal neuronal polarity specified by spatially localized mPar3/mPar6 and PI 3-kinase activity. *Cell* **112**, 63–75 (2003).
- Gassama-Diagne, A. *et al.* Phosphatidylinositol-3,4,5-trisphosphate regulates the formation of the basolateral plasma membrane in epithelial cells. *Nature Cell Biol.* **8**, 963–970 (2006).
- Weiner, O. D. *et al.* A PtdInsP<sub>3</sub>- and Rho GTPase-mediated positive feedback loop regulates neutrophil polarity. *Nature Cell Biol.* **4**, 509–513 (2002).
- Brandman, O. *et al.* Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science* **310**, 496–498 (2005).
- Irazoqui, J. E., Gladfelter, A. S. & Lew, D. J. Scaffold-mediated symmetry breaking by Cdc42p. *Nature Cell Biol.* **5**, 1062–1070 (2003).
- Marco, E. *et al.* Endocytosis optimizes the dynamic localization of membrane proteins that regulate cortical polarity. *Cell* **129**, 411–422 (2007).
- Ozbudak, E. M., Becskei, A. & van Oudenaarden, A. A system of counteracting feedback loops regulates Cdc42p activity during spontaneous cell polarization. *Dev. Cell* **9**, 565–571 (2005).
- Kozlov, M. M. & Mogilner, A. Model of polarization and bistability of cell fragments. *Biophys. J.* **93**, 3811–3819 (2007).
- Levin, S. A. in *Pattern Formation by Dynamic Systems and Pattern Recognition* (ed. Haken, H.) 210–222 (Springer, Berlin, 1979).
- Qian, H., Saffarian, S. & Elson, E. L. Concentration fluctuations in a mesoscopic oscillating chemical reaction system. *Proc. Natl Acad. Sci. USA* **99**, 10376–10381 (2002).
- Samoilov, M. S., Price, G. & Arkin, A. P. From fluctuations to phenotypes: the physiology of noise. *Sci. STKE* **2006**, re17 (2006).
- Suel, G. M. *et al.* Tunability and noise dependence in differentiation dynamics. *Science* **315**, 1716–1719 (2007).
- Elf, J. & Ehrenberg, M. Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Syst. Biol.* **1**, 230–236 (2004).
- Bose, I. *et al.* Assembly of scaffold-mediated complexes containing Cdc42p, the exchange factor Cdc24p, and the effector Cla4p required for cell cycle-regulated phosphorylation of Cdc24p. *J. Biol. Chem.* **276**, 7176–7186 (2001).
- Shimada, Y. *et al.* The nucleotide exchange factor Cdc24p may be regulated by auto-inhibition. *EMBO J.* **23**, 1051–1062 (2004).
- Chisari, M. *et al.* Shuttling of G protein subunits between the plasma membrane and intracellular membranes. *J. Biol. Chem.* **282**, 24092–24098 (2007).
- DerMardirossian, C. & Bokoch, G. M. GDIs: central regulatory molecules in Rho GTPase activation. *Trends Cell Biol.* **15**, 356–363 (2005).
- Bustelo, X. R., Sauzeau, V. & Berenjeno, I. M. GTP-binding proteins of the Rho/Rac family: regulation, effectors and functions *in vivo*. *Bioessays* **29**, 356–370 (2007).
- Ayscough, K. R. *et al.* High rates of actin filament turnover in budding yeast and roles for actin in establishment and maintenance of cell polarity revealed using the actin inhibitor latrunculin-A. *J. Cell Biol.* **137**, 399–416 (1997).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Altschuler, A. Artyukhin, T. Kurtz, C. Neuhauser, M. Rosen, R. Ranganathan, B. Shraiman, G. Süel and O. Weiner for their positive feedback. We additionally thank P. Crews for latrunculin A and R. Li for the yeast strain. This research was supported by an NIH grant (R01 GM071794), an NSF grant (DMS 0405084), the Endowed Scholars program at UT Southwestern Medical Center, and the Welch Foundation (I-1619, I-1644).

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to S.J.A. ([steven.altschuler@utsouthwestern.edu](mailto:steven.altschuler@utsouthwestern.edu)) or L.F.W. ([lani.wu@utsouthwestern.edu](mailto:lani.wu@utsouthwestern.edu)).

## METHODS

**Deterministic formulation of model.** The fraction  $h(t)$  of all molecules on the membrane changes in time  $t$  according to the ordinary differential equation:

$$\frac{dh}{dt} = k_{\text{on}}(1-h) + k_{\text{fb}}(1-h)h - k_{\text{off}}h$$

The density of membrane-bound molecules  $u(x, t)$  evolves in time according to the partial differential equation:

$$\frac{\partial u}{\partial t} = \frac{1}{2} D \nabla^2 u + k_{\text{on}}(1-h) + k_{\text{fb}}(1-h)u - k_{\text{off}}u$$

where  $\nabla^2$  is the diffusion operator on the plasma membrane.

**Stochastic formulation of model.** In the stochastic model we consider  $N$  molecules in a cell,  $n(t)$  of which are on the membrane at any given time. The molecules on the membrane have positions  $x_1(t), \dots, x_{n(t)}(t)$ , which undergo Brownian motion with diffusion coefficient  $D$ . The number of molecules on the membrane can also change randomly in three different ways: (1) A molecule on the membrane may dissociate itself from the membrane (rate constant  $k_{\text{off}}$ ); (2) a molecule on the membrane may 'recruit' some molecule in the cytosol to its location on the membrane (rate constant  $k_{\text{fb}}$ ); (3) a molecule from the cytosol may spontaneously associate itself to some random location on the membrane (rate constant  $k_{\text{on}}$ ).

The probability  $p_n(t)$  of  $n$  molecules appearing on the membrane at time  $t$  changes according to the master equation:  $\frac{dp_n(t)}{dt} = W_n - W_{n-1}$  where  $W_n = k_{\text{off}}(n+1)p_{n+1} - (k_{\text{on}} + k_{\text{fb}}n/N)(N-n)p_n$  while the number  $n(t)$  of molecules on the membrane evolves by the Markov process determined by the master equation. See Supplementary Information for a more detailed description of the stochastic model.

**Numerical simulations.** Whereas for the theoretical analysis the membrane may be a one-dimensional curve or a two-dimensional surface of arbitrary geometry, numerical simulations were carried out on one-dimensional circles for ease of visualization. Simulation results shown in the figures are based on model parameter ranges estimated from experimental measurements of polarized Cdc42 in budding yeast<sup>8,15</sup>. For a yeast cell of radius  $R \approx 5 \mu\text{m}$ , the plasma membrane diffusion constant of Cdc42 was estimated to be  $D = 0.001 (2\pi R)^2 \text{ min}^{-1} \approx 1.2 \mu\text{m}^2 \text{ min}^{-1} (= 5 \times 10^{-10} \text{ cm}^2 \text{ s}^{-1})$ . We assumed total Cdc42 numbers on the order of  $N \approx 10^3$  with plasma membrane fraction  $h_{\text{eq}} \approx 0.1$ . Fluorescence recovery after photobleaching (FRAP) analysis of

half-time recovery for polarized Cdc42 (as well as Cdc24, and Bem1) was on the order of  $\sim 5 \text{ s}$  (ref. 8); thus we estimated an off-rate of  $k_{\text{off}} = (\ln 2)/T_{1/2} \approx 9 \text{ min}^{-1}$ . A feedback strength of  $k_{\text{fb}} = 10 \text{ min}^{-1}$  was used to obtain an equilibrium plasma membrane fraction  $h_{\text{eq}} = (1 - k_{\text{off}}/k_{\text{fb}}) \approx 0.1$ . On-rate values  $k_{\text{on}}$  were varied as indicated in each figure legend. Clustering is predicted to occur for  $k_{\text{on}} \ll k_{\text{fb}}(h_{\text{eq}}(1 - h_{\text{eq}})N^2)^{-1} \approx 10^{-4} \text{ min}^{-1}$ , although simulations indicate that clustering can even occur for  $k_{\text{on}}$  less than  $10^{-3}$  (Fig. 3b and Supplementary Information). Polarization in simulations was determined by whether an interval of 10% of the membrane contained more than 50% of total membrane content; error bars represent standard errors from 4 replicates of 1,000 independent simulations (Fig. 3b, top panel).

**Release assay.** Cells from yeast strain RLY1948 (a gift from R. Li), which contain inducible GFP-Cdc42 expression under the Gal1/10 promoter, were grown logarithmically in synthetic complete (SC) medium without methionine<sup>8,15</sup>. To arrest cells in G1, cells were washed three times with autoclaved water, resuspended into YP medium supplemented with 2 mM methionine and 2% raffinose and distributed into four tubes. Cells were then cultured for 150 min (tube 1), 135 min (tube 2), 120 min (tube 3) and 90 min (tube 4). To induce expression of GFP-Cdc42, 2% galactose was added to each tube and cells grown for an additional 30 min (tube 1), 45 min (tube 2), 60 min (tube 3) and 90 min (tube 4). To release from G1 arrest, cells were washed three times with water and resuspended in methionine-free SC medium containing 2% glucose plus 150  $\mu\text{M}$  latrunculin A. The released cells were seeded in glass-bottomed microwell dishes (MatTek) and coated with concanavalin A for 10 min before imaging.

**Microscopy.** Live cell microscopy was performed on a Nikon epifluorescence microscope (Nikon TE-2000E2) equipped with a 100 plan apo objective, a cooled CCD camera (model COOLSNAP HQ; Photometrics Instruments) and a temperature control box set at 30 °C. Image acquisition was performed with Metamorph software (Universal Imaging Corp.) with  $1 \times 1$  binning. Images were taken in one focal plane at different locations in the dishes. For each condition, images were taken during a 30 min period.

**Image analysis.** Six images from each condition were selected for analysis yielding 145, 243, 211 and 228 individual cells, respectively. Using ImageJ, we manually identified cell regions and computed a background-subtracted, averaged GFP-Cdc42 intensity score. The presence of a Cdc42 polar cap in each cell was assessed (Supplementary Fig. 3), and probabilities of cap formation computed using custom Matlab software.

## LETTERS

# *Drosophila* RNAi screen identifies host genes important for influenza virus replication

Linhui Hao<sup>1,2\*</sup>, Akira Sakurai<sup>3\*†</sup>, Tokiko Watanabe<sup>3</sup>, Ericka Sorensen<sup>1</sup>, Chairul A. Nidom<sup>5,6</sup>, Michael A. Newton<sup>4</sup>, Paul Ahlquist<sup>1,2</sup> & Yoshihiro Kawaoka<sup>3,7,8,9</sup>

All viruses rely on host cell proteins and their associated mechanisms to complete the viral life cycle. Identifying the host molecules that participate in each step of virus replication could provide valuable new targets for antiviral therapy, but this goal may take several decades to achieve with conventional forward genetic screening methods and mammalian cell cultures. Here we describe a novel genome-wide RNA interference (RNAi) screen in *Drosophila*<sup>1</sup> that can be used to identify host genes important for influenza virus replication. After modifying influenza virus to allow infection of *Drosophila* cells and detection of influenza virus gene expression, we tested an RNAi library against 13,071 genes (90% of the *Drosophila* genome), identifying over 100 for which suppression in *Drosophila* cells significantly inhibited or stimulated reporter gene (*Renilla luciferase*) expression from an influenza-virus-derived vector. The relevance of these findings to influenza virus infection of mammalian cells is illustrated for a subset of the *Drosophila* genes identified; that is, for three implicated *Drosophila* genes, the corresponding human homologues *ATP6V0D1*, *COX6A1* and *NXF1* are shown to have key functions in the replication of H5N1 and H1N1 influenza A viruses, but not vesicular stomatitis virus or vaccinia virus, in human HEK 293 cells. Thus, we have demonstrated the feasibility of using genome-wide RNAi screens in *Drosophila* to identify previously unrecognized host proteins that are required for influenza virus replication. This could accelerate the development of new classes of antiviral drugs for chemoprophylaxis and treatment, which are urgently needed given the obstacles to rapid development of an effective vaccine against pandemic influenza and the probable emergence of strains resistant to available drugs.

Influenza, a highly contagious disease of birds and mammals, is caused by negative-strand RNA viruses of the family Orthomyxoviridae. Influenza outbreaks kill millions of people worldwide during pandemic years and hundreds of thousands during other years. Since their first lethal infection of humans in 1997, H5N1 influenza A viruses have spread throughout Asia and to Europe and Africa, posing a major risk for a new influenza pandemic<sup>2</sup>. To provide rational bases for improved treatment and control of influenza virus infection, we sought to advance understanding of viral infection mechanisms by elucidating previously unknown virus–host cell interactions. Many steps in the viral life cycle, including intracellular trafficking, gene expression, replication and virion assembly, depend on interactions with specific host cell gene products. Although most such host molecules remain elusive, emerging results indicate that

their identification and characterization can provide new insights into the mechanisms by which viruses complete their life cycle, and hence illuminate potentially valuable targets for prophylactic and therapeutic intervention<sup>3–5</sup>.

Systematic, genome-wide RNAi analysis offers an exciting tool to identify host genes that function in viral replication. Such analysis is facilitated by well-developed model systems such as *Drosophila*, the genome of which contains only ~14,000 genes, nearly all of which can be specifically targeted for high efficiency messenger RNA depletion by double-stranded RNA (dsRNA) libraries<sup>1</sup>. Because of its powerful genetics and conservation with vertebrates, *Drosophila* has been used to make numerous critical contributions to mammalian cell biology<sup>6–9</sup>. Thus, in principle, *Drosophila* RNAi studies could accelerate identification of host interactions essential for influenza virus replication.

Because *Drosophila* D-Mel2 cells do not express the human influenza virus receptor  $\alpha 2,6$ -linked sialic acid (Supplementary Fig. 1), we predicted that wild-type human influenza virus would not be able to infect them. Indeed, we did not detect viral protein expression by immunofluorescence assays in *Drosophila* D-Mel2 cells inoculated with influenza virus A/WSN/33 (WSN; H1N1) (data not shown). To bypass this block to wild-type influenza virus entry, we generated a genetically modified virus, Flu-VSV-G-GFP (FVG-G), in Madin-Darby canine kidney (MDCK) cells by replacing the receptor-binding haemagglutinin (HA) and neuraminidase (NA) genes with genes encoding vesicular stomatitis virus glycoprotein G (VSV-G) and enhanced green fluorescence protein (eGFP), respectively<sup>10,11</sup> (Fig. 1a). Because the envelopes of the resulting virions bear VSV-G, which mediates entry into mammalian, *Drosophila* and other cells<sup>12</sup>, FVG-G virions should readily infect *Drosophila* cells. Twenty-four hours after infection, GFP fluorescence was detected in FVG-G-infected *Drosophila* D-Mel2 cells (Supplementary Fig. 2). We also confirmed by real-time polymerase chain reaction (PCR) that influenza virus RNA replication occurred in *Drosophila* cells (Supplementary Fig. 3). However, *Drosophila* cells infected with FVG-G did not release detectable virions into the medium, as assayed by infectivity tests on MDCK cells and by electron microscopy (data not shown). This was due, at least in part, to failure of *Drosophila* cells to express some viral proteins required for virion assembly and infectivity (Supplementary Fig. 4). Thus, *Drosophila* cells can support influenza virus replication from post-entry to at least the protein expression phase of the viral life cycle. This span encompasses multiple other steps in the life cycle, including cytoplasmic release of

<sup>1</sup>Institute for Molecular Virology, <sup>2</sup>Howard Hughes Medical Institute, <sup>3</sup>Department of Pathobiological Sciences, and <sup>4</sup>Departments of Statistics and of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. <sup>5</sup>Faculty of Veterinary Medicine, and <sup>6</sup>Collaborating Research Center-Emerging & Reemerging Infectious Diseases, Tropical Disease Centre, Airlangga University, Surabaya 60115, Indonesia. <sup>7</sup>Division of Virology, Department of Microbiology and Immunology, and <sup>8</sup>International Research Center for Infectious Diseases, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. <sup>9</sup>Division of Zoonosis, Department of Microbiology and Infectious Diseases, Graduate School of Medicine, Kobe University, Kobe 650-0017, Japan. <sup>†</sup>Present address: First Department of Forensic Science, National Research Institute of Police Science, 6-3-1 Kashiwanoha, Kashiwa 277-0882, Japan.

\*These authors contributed equally to this work.

genomic RNA-containing viral ribonucleoprotein complexes (vRNPs), vRNP import into the nucleus, mRNA synthesis from the negative-strand viral RNA genome, mRNA export to the cytoplasm and translation.

For high-throughput, functional genomics analysis of influenza virus replication in *Drosophila* cells, we engineered Flu-VSV-G-R.Luc (FVG-R), in which VSV-G and *Renilla* luciferase genes replaced the viral HA and NA open reading frames (Fig. 1b). FVG-R virions were then used with an RNAi library (Ambion) against 13,071 *Drosophila* genes (~90% of all genes) to identify host genes affecting influenza-virus-directed *Renilla* luciferase expression (Fig. 1c). Two independent tests of the entire library were performed (Supplementary Table 1). For 176 genes for which dsRNAs inhibited FVG-R-directed luciferase expression in both replicates, repeated secondary tests using alternate dsRNAs to control for possible off-target effects confirmed the effects of 110 genes (Supplementary Tables 2 and 3). This confirmation rate is comparable to that in a *Drosophila* screen with a natural *Drosophila*-infecting virus<sup>5</sup>. Cell viability testing identified six genes with potentially significant

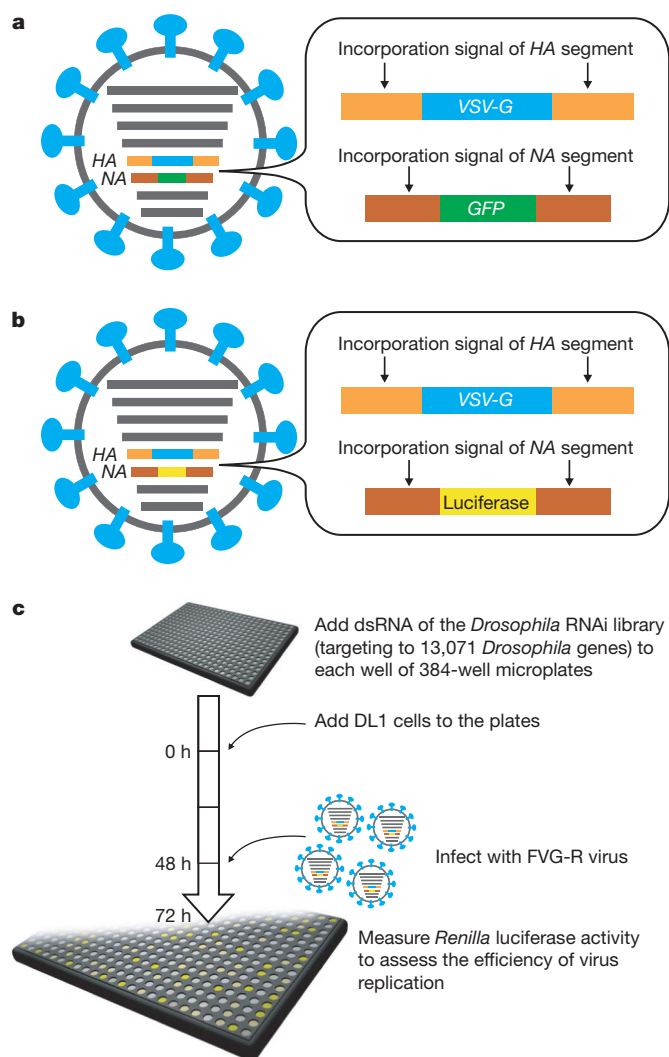
cytotoxic effects; these were excluded from further consideration (Supplementary Information and Supplementary Table 3). Secondary tests of candidate genes for which dsRNAs increased FVG-R-directed luciferase expression produced a much lower confirmation rate, suggesting a higher rate of off-target or other false-positive effects in this class (Supplementary Information and Supplementary Table 4).

Among the over 100 candidate genes found to be important for influenza virus replication in *Drosophila* cells, we selected the human homologues of several encoding components in host pathways/machineries that are known to be involved in the life cycle of influenza virus, for example, *ATP6V0D1* (endocytosis pathway), *COX6A1* (mitochondrial function) and *NXF1* (mRNA nuclear export machinery), for further analysis in mammalian cells to assess the relevance of our *Drosophila* results<sup>13–17</sup>. *ATP6V0D1* encodes subunit D of vacuolar (H<sup>+</sup>)-ATPase (V-ATPase), a proton pump that functions in the endocytosis pathway (that is, the acidification and fusion of intracellular compartments<sup>18</sup>). *COX6A1* encodes a subunit of cytochrome *c* oxidase (COX), an enzyme of the mitochondrial electron transport chain that catalyses electron transfer from cytochrome *c* to oxygen<sup>19</sup>. *NXF1* encodes a nuclear export factor critical for exporting most cellular mRNAs containing exon–exon junctions<sup>20,21</sup>.

As a first test for the possible contribution of these gene products to influenza virus replication in mammalian cells, we treated human HEK 293 cells twice at 24-h intervals with short interfering RNAs (siRNAs; siGENOME, Dharmacon) against the human homologue of each selected *Drosophila* gene. Twenty-four hours after the second siRNA treatment, the cells were infected with FVG-R virus and, two days later, *Renilla* luciferase activity was measured to assess viral replication and gene expression. siRNA against *ATP6V0D1* or *COX6A1* markedly decreased *Renilla* luciferase activity (Fig. 2a), but not cell viability (Supplementary Fig. 5a), suggesting that these genes have important roles in influenza virus replication in mammalian cells, as in *Drosophila* cells. Inhibition was not caused by off-target effects because, for each gene, each of four distinct siRNAs inhibited FVG-R-directed expression of *Renilla* luciferase (Supplementary Table 6). Because *COX6A1* encodes a subunit of mitochondrial electron transport chain complex IV, COX, we used specific inhibitors to test whether in HEK 293 cells influenza virus also required other complexes in this chain (Fig. 2c). Inhibitors of complexes III, IV and V selectively inhibited FVG-R-directed *Renilla* luciferase expression by 50- to 100-fold, whereas complex I and II inhibitors had little or no effect. Thus, in mammalian cells, influenza virus depends on multiple late stages but not early stages in the mitochondrial electron transport chain.

Treatment for four days with siRNA against *NXF1* decreased mammalian cell viability (data not shown), as predicted by the critical role of *NXF1* in general host cell metabolism. Accordingly, the total incubation time with siRNA against *NXF1* was shortened to 36 h by transfecting cells with the siRNA twice at a 12-h interval, infecting with FVG-R virus 12 h later, and assaying for *Renilla* luciferase at 12-h post-infection. Under these conditions, cell viability was not detectably affected (Supplementary Fig. 5b) whereas *Renilla* luciferase activity was reduced by nearly fivefold (Fig. 2b). Whereas recent results indicated that influenza virus protein NS1 binds to *NXF1* to inhibit host mRNA export<sup>17</sup>, these results imply that influenza virus RNAs and/or proteins are transported by an *NXF1*-dependent pathway (see also Supplementary Information).

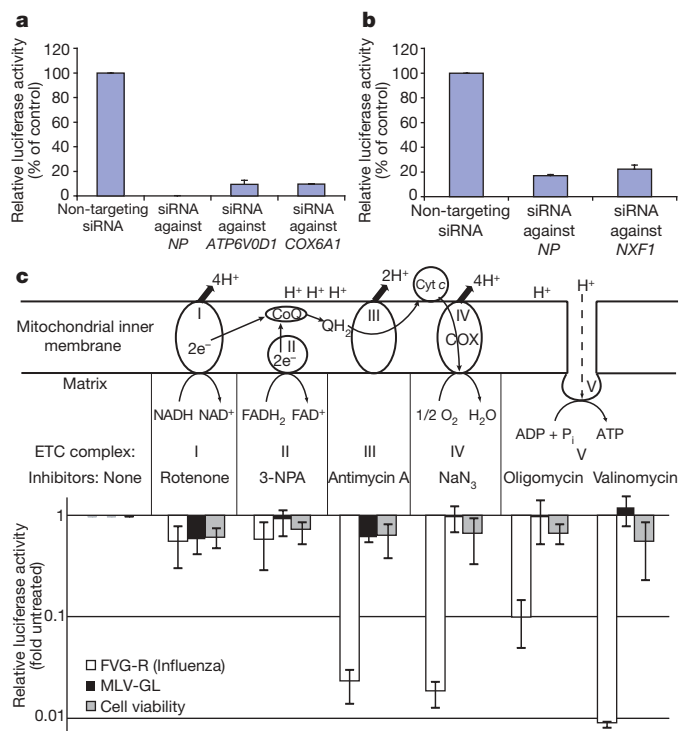
To test the effects of these genes on authentic influenza viruses, we infected siRNA-treated HEK 293 cells with WSN virus or H5N1 influenza A/Indonesia/7/05 (Indonesia 7; isolated from a patient) or with VSV or vaccinia virus as controls. Progeny viruses were collected from the medium at 24 h (Indonesia 7, VSV or vaccinia virus) or 48 h (WSN) post-infection and were titrated. Depleting *ATP6V0D1* and *COX6A1* did not affect VSV or vaccinia virus replication, but decreased the WSN and Indonesia 7 virus yields by ~10-fold or more (Fig. 3a). Thus, *ATP6V0D1* and *COX6A1* are required for replication



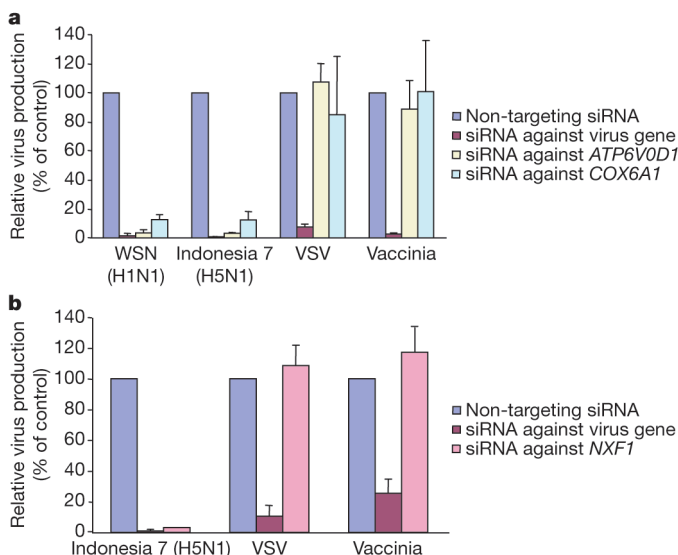
**Figure 1 | Overview of genome-wide RNAi screen to identify host factors involved in influenza virus replication in *Drosophila* cells.** **a, b**, Schematic diagrams showing recombinant influenza viruses. Shown are FVG-G, in which genes encoding the HA and NA proteins were replaced with the VSV-G and *eGFP* genes, respectively (**a**), and FVG-R, in which the genes encoding the HA and NA were replaced with the VSV-G and *Renilla* luciferase genes, respectively (**b**). **c**, Schematic diagram of the systematic analysis of host genes affecting influenza virus replication and gene expression in *Drosophila* cells. Experimental details are given in Methods.

of influenza viruses but not VSV and vaccinia virus. Depleting the *NXF1* nuclear export factor, again using an accelerated 36-h time line, reduced Indonesia 7 virus titres by 20-fold (Fig. 3b), indicating that *NXF1* has a critical role in H5N1 influenza virus replication. Unfortunately, because WSN grows in HEK 293 cells more slowly than does Indonesia 7 virus, the WSN virus yield at 12-h post-infection was insufficient to test for an effect of *NXF1* depletion (data not shown). In comparison to Indonesia 7 virus, *NXF1* depletion had no effect on VSV or vaccinia virus yields (Fig. 3b). Moreover, siRNA against *NXF1* inhibited FVG-G, but not adenovirus, which, like influenza virus, depends on nuclear steps for genome transcription and replication (Supplementary Fig. 6). Thus, multiple genes identified in *Drosophila* cells correspond to important, selective host factors for influenza virus replication in mammalian cells.

The method that we have established, using systematic analysis of a *Drosophila* RNAi library with confirmation in mammalian cells, can be used to identify host gene products that affect influenza virus replication. This utility was demonstrated in experiments with authentic influenza viruses, including an H5N1 virus isolated from a human patient. One of the candidate genes tested in mammalian cells, *COX6A1*, which encodes a subunit of COX, acts as a critical enzyme in cytochrome-*c*-dependent electron transport in mitochondria. Influenza virus PB2 polymerase has a mitochondrial targeting signal<sup>14</sup>, and PB1-F2, the second protein encoded by the influenza *PB1* gene, was reported to localize to inner and outer mitochondrial membranes and to delay influenza virus clearance by host antiviral responses<sup>15</sup>. Thus, *COX6A1* may be involved in PB2- and/or PB1-F2-mediated functions in mitochondria. Additionally, influenza vRNP



**Figure 2 | Effect of selected siRNAs and inhibitors on *Renilla* luciferase expression in FVG-R-infected human cells.** **a–c**, *Renilla* luciferase activity was measured in FVG-R-infected HEK 293 cells treated with siRNAs against *ATP6V0D1* and *COX6A1* (**a**), *NXF1* (**b**) or the indicated mitochondrial electron transport chain inhibitors (**c**). Inhibitors of complexes III, IV and V inhibited FVG-R-directed *Renilla* luciferase expression significantly, whereas complex I and II inhibitors had little or no effect. In contrast, the inhibitors had no significant effects on cell viability and *Gaussia* luciferase expression of a murine leukaemia virus derivative (MLV-GL) that, like FVG-R, depended on VSV G-mediated entry. All experiments were conducted three times in duplicate, with the results reported as mean  $\pm$  s.d. 3-NPA, 3-nitropropionic acid.



**Figure 3 | Effect of siRNAs against selected genes on the replication of influenza viruses, VSV or vaccinia virus in human HEK 293 cells.** **a, b**, The titres of influenza viruses (WSN and Indonesia 7), VSV and vaccinia virus in HEK 293 cells treated with siRNA against *ATP6V0D1* and *COX6A1* (**a**) or against *NXF1* (**b**) are shown. Experimental details are given in Methods and Supplementary Methods. All experiments were conducted three times, with the results reported as mean  $\pm$  s.d.

export from the nucleus requires caspase 3 activation<sup>22</sup>, which can be induced by cytochrome *c* release from mitochondria. Moreover, human BCL2 inhibits both cytochrome *c* release and influenza vRNP export<sup>23</sup>, and COX function and changes in mitochondrial membrane potential have been linked to caspase activation<sup>24</sup>. These pathways may underlie our further findings (Fig. 2c) that influenza virus replication and expression in mammalian cells are strongly and selectively modulated by compounds that inhibit COX, cytochrome *c*-linked electron transport chain complex III, and normal ion transport across mitochondrial membranes (oligomycin and valinomycin). Potential roles of the other two host genes confirmed here in mammalian cells, *ATP6V0D1* and *NXF1*, in influenza virus replication are discussed further in the Supplementary Information.

The above results with multiple, diverse genes, including *ATP6V0D1*, *NXF1* and *COX6A1* as well as mitochondrial electron transport complexes III and V, demonstrate the feasibility and value of using *Drosophila* RNAi screening to identify previously unknown host factors with important and potentially unsuspected roles in influenza virus replication. Simultaneously, the genome-wide results from our *Drosophila* RNAi screen provide more than 100 additional candidate genes (many with unknown functions) to be tested in mammalian cells. We suggest that the same strategy could be applied to identify previously unknown host factors involved in the replication of other viruses, whenever at least a portion of their replication cycle is supported by *Drosophila* cells.

## METHODS SUMMARY

**Cells and viruses.** MDCK cells, HEK 293 cells, baby hamster kidney (BHK) cells, D-Mel2 cells and DL1 cells were maintained as described in Methods. WSN, FVG-G, FVG-R and Indonesia 7 viruses were generated as described previously<sup>10</sup> and propagated in MDCK cells. VSV and vaccinia virus were grown in BHK and CV-1 cells, respectively. *Gaussia*-luciferase-expressing murine leukaemia virus (MLV-GL) and the GFP-expressing adenovirus were provided by J. Bruce and R. Kalejta<sup>25</sup>, respectively.

***Drosophila* RNAi library analysis.** Double-stranded RNA of the *Drosophila* RNAi library (targeting 13,071 *Drosophila* genes) and DL1 cells were added to each well of 384-well microplates. After two days of incubation, cells were infected with the FVG-R virus. At one day post-infection, *Renilla* luciferase activity was measured as described below. Two independent analyses of the entire library were performed.

**siRNA treatment of mammalian cells.** HEK 293 cells were transfected with siRNA by TransIT-TKO (Mirus). Cells were incubated for 12 h in experiments with *NXF1* or for 24 h with other genes, were transfected again under the same conditions, and inoculated with virus after 12 h for *NXF1* or after 24 h for other genes. After a further 12 h (for *NXF1*) or 48 h (other genes), the cells were harvested for the indicated analyses.

**Renilla luciferase and cell viability assays.** *Renilla* luciferase activity and cell viability were measured with established *Renilla* luciferase and CellTiter-Glo assay systems (Promega) according to the manufacturer's instructions; signals were read with a GLOMAX 96 microplate luminometer.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 15 October 2007; accepted 5 June 2008.**

**Published online 9 July 2008.**

- Kuttenkeuler, D. & Boutros, M. Genome-wide RNAi as a route to gene function in *Drosophila*. *Brief. Funct. Genomic. Proteomic.* **3**, 168–176 (2004).
- Webster, R. G. & Govorkova, E. A. H5N1 influenza — continuing evolution and spread. *N. Engl. J. Med.* **355**, 2174–2177 (2006).
- Kushner, D. B. *et al.* Systematic, genome-wide identification of host genes affecting replication of a positive-strand RNA virus. *Proc. Natl Acad. Sci. USA* **100**, 15764–15769 (2003).
- Ludwig, S., Planz, O., Pleschka, S. & Wolff, T. Influenza-virus-induced signaling cascades: targets for antiviral therapy? *Trends Mol. Med.* **9**, 46–52 (2003).
- Cherry, S. *et al.* Genome-wide RNAi screen reveals a specific sensitivity of IRES-containing RNA viruses to host translation inhibition. *Genes Dev.* **19**, 445–452 (2005).
- Hahn, H., Wojnowski, L., Miller, G. & Zimmer, A. The patched signaling pathway in tumorigenesis and development: lessons from animal models. *J. Mol. Med.* **77**, 459–468 (1999).
- Dearolf, C. R. JAKs and STATs in invertebrate model organisms. *Cell. Mol. Life Sci.* **55**, 1578–1584 (1999).
- Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M. & Hoffmann, J. A. The dorsoventral regulatory gene cassette *spatzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973–983 (1996).
- Takeda, K., Kaisho, T. & Akira, S. Toll-like receptors. *Annu. Rev. Immunol.* **21**, 335–376 (2003).
- Neumann, G. *et al.* Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl Acad. Sci. USA* **96**, 9345–9350 (1999).
- Watanabe, T., Watanabe, S., Noda, T., Fujii, Y. & Kawaoka, Y. Exploitation of nucleic acid packaging signals to generate a novel influenza virus-based vector stably expressing two foreign genes. *J. Virol.* **77**, 10575–10583 (2003).
- Wyers, F., Richard-Molard, C., Blondel, D. & Dezelee, S. Vesicular stomatitis virus growth in *Drosophila melanogaster* cells: G protein deficiency. *J. Virol.* **33**, 411–422 (1980).
- Perez, L. & Carrasco, L. Involvement of the vacuolar H(+) -ATPase in animal virus entry. *J. Gen. Virol.* **75**, 2595–2606 (1994).
- Carr, S. M., Carnero, E., Garcia-Sastre, A., Brownlee, G. G. & Fodor, E. Characterization of a mitochondrial-targeting signal in the PB2 protein of influenza viruses. *Virology* **344**, 492–508 (2006).
- Zamarin, D., Ortigoza, M. B. & Palese, P. Influenza A virus PB1-F2 protein contributes to viral pathogenesis in mice. *J. Virol.* **80**, 7976–7983 (2006).
- Momose, F. *et al.* Cellular splicing factor RAF-2p48/NPI-5/BAT1/UAP56 interacts with the influenza virus nucleoprotein and enhances viral RNA synthesis. *J. Virol.* **75**, 1899–1908 (2001).
- Satterly, N. *et al.* Influenza virus targets the mRNA export machinery and the nuclear pore complex. *Proc. Natl Acad. Sci. USA* **104**, 1853–1858 (2007).
- Stevens, T. H. & Forgac, M. Structure, function and regulation of the vacuolar (H+) -ATPase. *Annu. Rev. Cell Dev. Biol.* **13**, 779–808 (1997).
- Gnaiger, E., Lassnig, B., Kuznetsov, A., Rieger, G. & Margreiter, R. Mitochondrial oxygen affinity, respiratory flux control and excess capacity of cytochrome c oxidase. *J. Exp. Biol.* **201**, 1129–1139 (1998).
- Cullen, B. R. Nuclear mRNA export: insights from virology. *Trends Biochem. Sci.* **28**, 419–424 (2003).
- Reed, R. & Cheng, H. TREX, SR proteins and export of mRNA. *Curr. Opin. Cell Biol.* **17**, 269–273 (2005).
- Wurzer, W. J. *et al.* Caspase 3 activation is essential for efficient influenza virus propagation. *EMBO J.* **22**, 2717–2728 (2003).
- Hinshaw, V. S., Olsen, C. W., Dybdahl-Sissoko, N. & Evans, D. Apoptosis: a mechanism of cell killing by influenza A and B viruses. *J. Virol.* **68**, 3667–3673 (1994).
- Kadenbach, B., Arnold, S., Lee, I. & Huttemann, M. The possible role of cytochrome c oxidase in stress-induced apoptosis and degenerative diseases. *Biochim. Biophys. Acta* **1655**, 400–408 (2004).
- He, T. C. *et al.* A simplified system for generating recombinant adenoviruses. *Proc. Natl Acad. Sci. USA* **95**, 2509–2514 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. McGregor and K. Wells for technical assistance, M. Hatta, S. Yamada and M. Ito for an H5N1 influenza virus, S. Watanabe and Y. Hatta for FVG-R and FVG-G viruses, R. Kalejta and J. Bruce for adenovirus and murine leukaemia virus vectors, T. Noda for electron microscopy, J. Gilbert for editing the manuscript and Y. Kawaoka for illustrations. This work was supported, in part, by US National Institute of Allergy and Infectious Diseases Public Health Service research grants, by the US National Institute of General Medical Sciences grant GM35072, by a grant-in-aid for Specially Promoted Research and by a contract research fund for the Program of Funding Research Centers for Emerging and Reemerging Infectious Diseases from the Ministry of Education, Culture, Sports, Science and Technology, and by grants-in-aid from the Ministry of Health, Labor, Welfare of Japan. P.A. is a Howard Hughes Medical Institute investigator.

**Author Contributions** L.H., A.S., P.A. and Y.K. designed research; L.H., A.S., T.W. and C.A.N. performed research; L.H., A.S., T.W., E.S., M.A.N., P.A. and Y.K. analysed data; and L.H., A.S., T.W., P.A. and Y.K. wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to Y.K. ([kawaokay@svm.vetmed.wisc.edu](mailto:kawaokay@svm.vetmed.wisc.edu)) or P.A. ([ahlquist@wisc.edu](mailto:ahlquist@wisc.edu)).

## METHODS

**Cells and viruses.** MDCK cells, HEK 293 cells and BHK cells were maintained in minimum essential medium containing 5% fetal calf serum and antibiotics at 37 °C in 5% CO<sub>2</sub>. D-Mel2 cells were maintained in *Drosophila*-SFM (GIBCO/Invitrogen) at 28 °C. DL1 cells were maintained in Schneider's *Drosophila* (SD) medium containing 10% FBS at 28 °C.

WSN, FVG-G, FVG-R and Indonesia 7 viruses were generated by a plasmid-based reverse genetics system<sup>10</sup> and were grown and titrated in MDCK cells. MLV-GL was produced by replacing the CD4 open reading frame in pCMMP-CD4-eGFP<sup>26</sup> with that of *Gaussia* luciferase (J. Bruce, unpublished observations). All experiments with Indonesia 7 virus were conducted in a bio-safety level 3 containment laboratory approved for such use by the Centers for Disease Control and Prevention and the US Department of Agriculture.

***Drosophila* RNAi library analysis.** The *Drosophila* RNAi library (Ambion) contained 13,071 individual dsRNAs, each designed to specifically target a single *Drosophila* gene. Five microlitres of dsRNA (40 ng µl<sup>-1</sup>) were added to each well of 384-well plates, after which 2 × 10<sup>4</sup> DL1 cells in 10 µl of SD medium were added to each well and incubated with the dsRNA at 28 °C for 60 min. Twenty microlitres of SD medium containing 20% FBS was then added to each well after incubation. Cells were treated with dsRNA for two days at 28 °C and were then inoculated with an amount of FVG-R virus corresponding to a multiplicity of infection of 10 for MDCK cells, and were then transferred to 33 °C. At one day post-infection, *Renilla* luciferase activity was measured as described below. Two independent analyses of the entire library were performed.

**siRNA treatment of mammalian cells.** The non-targeted siRNA used was siCONTROL 1 (Dharmacon). Duplex siRNAs against candidate genes were obtained from siGENOME (Dharmacon, see Supplementary Table 7 for sequences); the effects of siRNA were evaluated by reverse transcription followed by PCR (RT-PCR; Supplementary Fig. 7). The sequences of siRNA against the NP gene of influenza virus (GGAUCUUAUUUCCUUCGAGUU)<sup>27</sup> and the E3L gene of vaccinia virus (AAUAUCGUCGGAGCUGUACAC)<sup>28</sup> were reported previously. The sequence of siRNA against the gene encoding the L protein of VSV (CGAGUUAUCCAGCAAUCAUUU) was designed using BLOCK-iT RNAi designer (Invitrogen). A HEK 293 cell suspension was seeded into the wells of a 24-well plate (2.0 × 10<sup>4</sup> cells per well), incubated for 1 h and transfected with siRNA (to a final concentration 10 nM; Dharmacon) using TransIT-TKO (Mirus), according to the manufacturer's instructions.

**Inhibitor treatment of mammalian cells.** Mitochondrial electron transport chain inhibitors were used at the following final concentrations: rotenone (0.5 µM), 3-nitropropionic acid (1 mM), antimycin A (25 µM), NaN<sub>3</sub> (50 mM), oligomycin (25 µM) and valinomycin (0.9 µM). HEK 293 cells were treated with the inhibitors for 5 h before being infected with FVG-R or MLV-GL, incubated 18 h, and assayed for *Renilla* or *Gaussia* luciferase, respectively. Cell viability was assayed in parallel by a firefly-luciferase-mediated ATP assay (CellTiter-Glo, Promega).

26. Bruce, J. W., Bradley, K. A., Ahlquist, P. & Young, J. A. Isolation of cell lines that show novel, murine leukemia virus-specific blocks to early steps of retroviral replication. *J. Virol.* **79**, 12969–12978 (2005).

27. Ge, Q. *et al.* RNA interference of influenza virus production by directly targeting mRNA for degradation and indirectly inhibiting all viral RNA transcription. *Proc. Natl Acad. Sci. USA* **100**, 2718–2723 (2003).

28. Dave, R. S. *et al.* siRNA targeting vaccinia virus double-stranded RNA binding protein [E3L] exerts potent antiviral effects. *Virology* **348**, 489–497 (2006).

## LETTERS

# Transient FTY720 treatment promotes immune-mediated clearance of a chronic viral infection

Mary Premenko-Lanier<sup>1</sup>, Nelson B. Moseley<sup>1</sup>, Sarah T. Pruett<sup>2</sup>, Pablo A. Romagnoli<sup>1</sup> & John D. Altman<sup>1</sup>

For a wide variety of microbial pathogens, the outcome of the infection is indeterminate. In some individuals the microbe is cleared, but in others it establishes a chronic infection, and the factors that tip this balance are often unknown. In a widely used model of chronic viral infection, C57BL/6 mice clear the Armstrong strain of lymphocytic choriomeningitis virus (LCMV), but the clone 13 strain persists<sup>1,2</sup>. Here we show that the Armstrong strain induces a profound lymphopenia at days 1–3 after infection, but the clone 13 strain does not. If we transiently augment lymphopenia by treating the clone-13-infected mice with the drug FTY720 at days 0–2 after infection, the mice successfully clear the infection by day 30. Clearance does not occur when CD4 T cells are absent at the time of treatment, indicating that the drug is not exerting direct antiviral effects. Notably, FTY720 treatment of an already established persistent infection also leads to viral clearance. In both models, FTY720 treatment preserves or augments LCMV-specific CD4 and CD8 T-cell responses, a result that is counter-intuitive because FTY720 is generally regarded as a new immunosuppressive agent<sup>3</sup>. Because FTY720 targets host pathways that are completely evolutionarily conserved, our results may be translatable into new immunotherapies for the treatment of chronic microbial infections in humans.

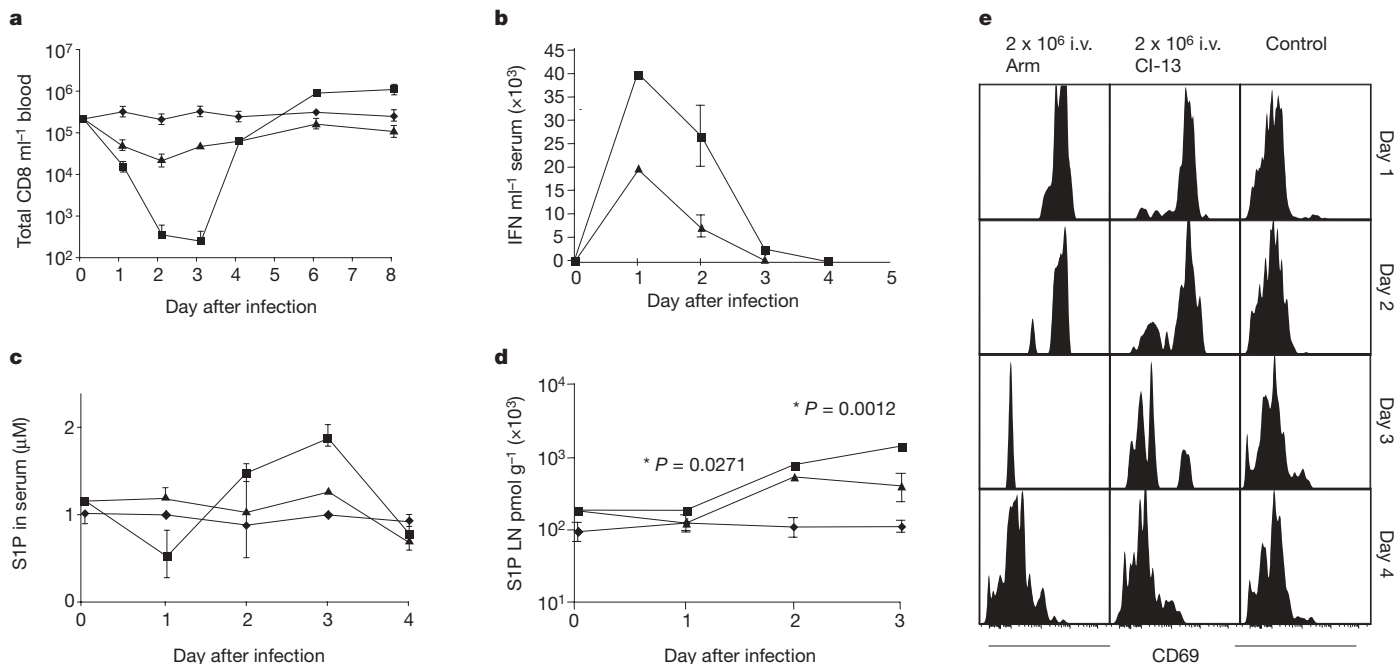
High-level persistent infections cause a variety of immune system dysfunctions<sup>4–7</sup>, and the initial focus of our study was to assess their effects on bystander naive T cells. For these experiments, we used the established LCMV model and infected mice with either the Armstrong or the clone 13 strain of LCMV. In the course of these experiments we observed that Armstrong infection, as expected<sup>8</sup>, induced an early and profound transient decrease in the total number of lymphocytes circulating in the blood (lymphopenia); notably, clone 13 infection induced lymphopenia that was much less profound (Fig. 1a and Supplementary Fig. 1a, b). Sequestration contributed significantly to the Armstrong-induced lymphopenia because a sentinel population of adoptively transferred F5 T-cell-receptor transgenic cells, specific for an influenza epitope not found in LCMV<sup>9</sup>, disappeared from the blood 1–3 days after infection but re-appeared between days 3 to 4 after infection (Supplementary Fig. 2a). In contrast, the total number of F5 cells in the blood and lymphoid organs remained constant, demonstrating that the blood lymphopenia was not due to global lymphocyte death in the Armstrong-infected mice (Supplementary Fig. 2b). During the period of lymphopenia, lymphocytes were quantified in secondary lymphoid tissues (Supplementary Fig. 3). Lymphocytes were observed to accumulate in the lymph nodes after both Armstrong and clone 13 infections, but the quantities were greater in the Armstrong-infected mice. In general, the total number of CD4 and CD8 T cells, and CD19 B cells, were decreased in the spleens of mice infected with both virus strains.

Sequestration of cells in lymph nodes is regulated at the level of both entry and exit, and the exit pathway is regulated by a G-protein-coupled receptor for sphingosine-1-phosphate (S1P), called S1P<sub>1</sub> (ref. 10). Although high quality reagents for directly measuring S1P<sub>1</sub> levels are not readily available, CD69 has recently been reported to downregulate S1P<sub>1</sub>, and upregulation of CD69 is induced by type I interferon (IFN-I)<sup>8</sup>. Seeking an explanation for the reduced lymphopenia after clone 13 infection, we performed daily measurements of the amounts of IFN-I and S1P in the serum of clone-13- and Armstrong-infected mice and uninfected controls; at the same time points, we also measured CD69 expression on lymphocytes. Compared to clone-13-infected mice, mice infected with the Armstrong strain had slightly higher serum levels of IFN-I that were significant on day 1 after infection (Fig. 1b,  $P = 0.0001$ ), slightly lower amounts of serum S1P (Fig. 1c,  $P = 0.0473$ , day 1), and slightly higher amounts of S1P in their lymph nodes (Fig. 1d,  $P = 0.0271$ , day 1, and  $P = 0.0012$ , day 3). However, despite the differences in IFN-I levels, there was no difference in the frequency of cells that carry the CD69 antigen in the blood, nor a difference in the levels of CD69 expression on CD8 T cells (Fig. 1e), CD4, or CD19 cells (data not shown). Therefore, we still lack an understanding of the causes of the difference in lymphopenia after infection with the two strains of LCMV.

The difference in lymphopenia after infection with Armstrong and clone 13 strains raises the question as to whether the immune system would be able to clear a clone 13 infection if we could artificially induce a more profound but still transient lymphopenia. To force the sequestration of lymphocytes we chose the immunomodulatory pro-drug FTY720, well known for its ability to sequester lymphocytes in lymph nodes and Peyer's patches<sup>11</sup>. When phosphorylated, FTY720-P is an S1P analogue capable of binding to four of the five S1P receptors. FTY720 has been tested in human clinical trials for the treatment of relapsing multiple sclerosis<sup>12</sup> and the prevention of kidney transplant rejection<sup>13</sup>, and it is widely regarded as a new class of immunosuppressive agent. In contrast to traditional immunosuppressive drugs, FTY720 is thought to work by manipulating lymphocyte location and has been shown to have no direct effect on the activation of T cells<sup>14</sup>.

To force the early sequestration of naive T cells during clone 13 infection, FTY720 was administered to mice intravenously at doses ranging from  $0.8 \mu\text{g kg}^{-1}$  to  $1 \text{ mg kg}^{-1}$ . In an attempt to mimic the transient lymphopenia seen after infection with the Armstrong strain, we chose a treatment regimen in which FTY720 was administered once a day for 3 days beginning at the time of infection. At low FTY720 doses, T cells were sequestered for 3 days after infection, but returned to the circulation by day 4; at higher drug doses, they remained sequestered for at least an extra day (Supplementary Fig. 4a, b). Furthermore, low doses of FTY720 caused a significant reduction in clone 13 serum titres at 8 days after infection, but the highest dose ( $1 \text{ mg kg}^{-1}$ ) did not (Supplementary Fig. 4c). On the basis of

<sup>1</sup>Emory Vaccine Center and Department of Microbiology and Immunology; Yerkes National Primate Research Center and Emory University School of Medicine, 954 Gatewood Road, Atlanta, Georgia 30329, USA. <sup>2</sup>Division of Psychobiology, Yerkes National Primate Research Center, 954 Gatewood Road, Atlanta, Georgia 30329, USA.



**Figure 1 | LCMV Armstrong but not clone 13 induces profound lymphopenia at days 2 and 3 after infection.** Wild-type mice were infected intravenously (i.v.) at  $2 \times 10^6$  p.f.u. with LCMV Armstrong strain (filled squares; Arm), clone 13 strain (filled triangles; Cl-13), or were uninfected (filled diamonds) and measurements were taken on successive days after infection. **a**, The concentration of lymphocytes (CD8 T cells) per ml of blood

is shown. **b**, IFN- $\gamma$  amounts in the serum; IFN- $\gamma$  was measured using a standard vesicular stomatitis virus protection assay using the L929 fibroblast cell line<sup>24</sup>. **c**, **d**, S1P concentrations in serum (**c**) and in lymph nodes (LN, **d**); S1P concentrations were measured by liquid chromatography tandem mass spectrometry;  $n = 3$  per group<sup>25</sup>. **e**, CD69 expression on CD8 T cells in the blood. All data are representative of 2–3 experiments. Error bars are s.d.

these results, we chose to use a dose of  $4 \mu\text{g kg}^{-1}$  for all subsequent experiments. Although this dose causes lymphopenia in clone-13-infected mice, it does not in uninfected controls (data not shown), suggesting that there is some synergy between the drug and the infected state. Also at this dose, FTY720 does not alter the concentrations of S1P in the serum or lymph nodes, the IFN- $\gamma$  amount in serum or the expression of CD69 on lymphocytes (data not shown). FTY720 causes a transient and statistically significant increase in the numbers of T and B cells in the mesenteric lymph nodes in clone-13-infected mice ( $P = 0.0267$ ); drug-associated increases in the numbers of lymphocytes are also seen in the peripheral lymph nodes, but these are not statistically significant (Supplementary Fig. 5a). In experiments using adoptively transferred LCMV-specific P14 T-cell-receptor transgenic cells, FTY720 treatment induced a significant increase in the numbers of P14 cells in the lymph nodes at days 3 and 4 after infection, consistent with the hypothesis that sequestration accelerates priming of the adaptive CD8<sup>+</sup> T-cell response (Fig. 2a).

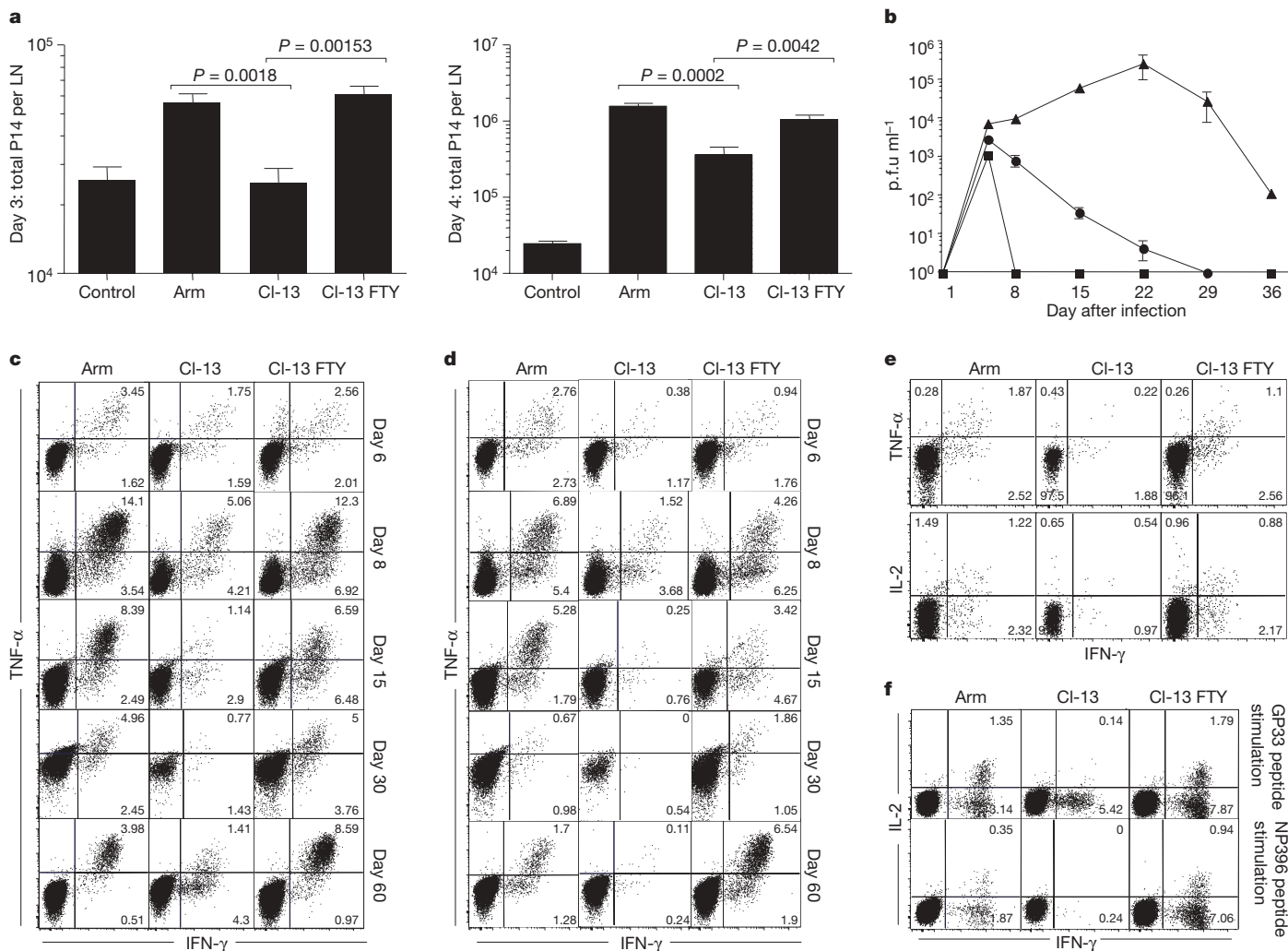
The data in Supplementary Fig. 4c prompted us to test more extensively the efficacy of FTY720 treatment of clone-13-infected mice. We focused on two measures of efficacy: viral titres in the serum, brain and kidneys, and T-cell function. Beginning as soon as 5 days after infection, FTY720-treated mice had significantly lower clone 13 titres in the serum than untreated mice ( $P = 0.0066$ ; Fig. 2b), and the drug-induced control of the virus continued until it was cleared by day 29 after infection. Viral titres in the kidneys and brain of the clone-13-infected, FTY720-treated mice are undetectable at day 8 and beyond, in contrast to the untreated clone-13-infected mice that had high viral titres in the kidney and brain up to at least day 60 after infection (data not shown).

T-cell exhaustion is a hallmark of chronic clone 13 infection<sup>7,15,16</sup>. In mice infected with clone 13, cells specific for the dominant epitope NP396 are deleted and the T cells specific for the co-dominant GP33 epitope become dysfunctional<sup>7</sup>. We found that mice treated with FTY720 at the time of clone 13 infection have functional CD8 T cells specific for the LCMV epitopes GP33 and NP396 throughout the

course of the study (Fig. 2c, d). We also analysed the LCMV-specific CD4 T-cell response and found that FTY720-treated but not untreated mice retained GP61-specific functional T cells (Fig. 2e). Furthermore, PD-1 expression on CD8 T cells is tightly linked to the presence of antigen (and T-cell exhaustion in the chronic phase of infection)<sup>17</sup>, and by day 15 after clone 13 infection, PD-1 amounts are lower on LCMV-specific cells in FTY720-treated but not untreated mice (Supplementary Fig. 6); this result correlates with the control of the infection. Neutralizing antibodies were measured on day 60 after infection using a simple plaque reduction assay, and FTY720 treatment did not alter their titres (data not shown). Finally, mice given FTY720 at the time of infection form a stable memory population, with fully functional CD8 T cells that are able to produce interleukin (IL)-2 on *in vitro* stimulation with the dominant epitopes NP396 and GP33 (Fig. 2f). Therefore, FTY720 treatment at the time of clone 13 infection leads to clearance of the virus and preservation of LCMV-specific T-cell responses.

We next investigated whether short-term FTY720 treatment could clear an already established clone 13 infection. Mice were infected with clone 13 and treated with FTY720 at  $4 \mu\text{g kg}^{-1}$  on days 30, 31 and 32 after infection. By day 30 after treatment—60 days after infection—the virus was cleared from the kidneys and brain (Fig. 3a;  $P = 0.0008$  unpaired *t*-test), organs that harbour virus for the lifetime of the untreated mouse. Within 2 weeks of FTY720 treatment, NP396 CD8 T-cell specific peptide responses and GP61 CD4 T-cell specific responses were detectable in the treated mice but not in the untreated controls (Fig. 3b); GP33-specific responses were detected in untreated mice but were significantly enhanced after treatment, as detected by both cytokine production (Fig. 3b) and tetramer staining (Fig. 3c). At 4 weeks after treatment, GP33-tetramer-positive cells in lymph nodes had downregulated expression of PD-1 and upregulated expression of CD62L, indicating a transition from effector to memory T cells (Fig. 3d).

The expansion of GP33-specific cells after FTY720 treatment during the persistent phase of infection was also seen in mice



**Figure 2 | Low-dose short-term FTY720 treatment of clone-13-infected mice prevents the establishment of a persistent LCMV infection.** Wild-type mice were infected intravenously with  $2 \times 10^6$  p.f.u. of LCMV Armstrong (Arm) or clone 13 (CI-13). FTY720 was administered intravenously at  $4 \mu\text{g kg}^{-1}$  on days 0, 1 and 2 after infection. For **a** only,  $1 \times 10^6$  P14 cells were transferred 48 h before infection. **a**, Total P14 cells in lymph nodes (LN) on days 3 (left) and 4 (right) after infection. **b**, Viral titres in serum were determined using a standard plaque assay from mice treated with Armstrong (squares), clone 13 (triangles) and clone 13 plus FTY720 (circles). Unpaired *t*-tests comparing clone-13-infected mice with or without FTY720 treatment were as follows:  $P = 0.0066$  day 5,  $P = 0.0001$  day 8,

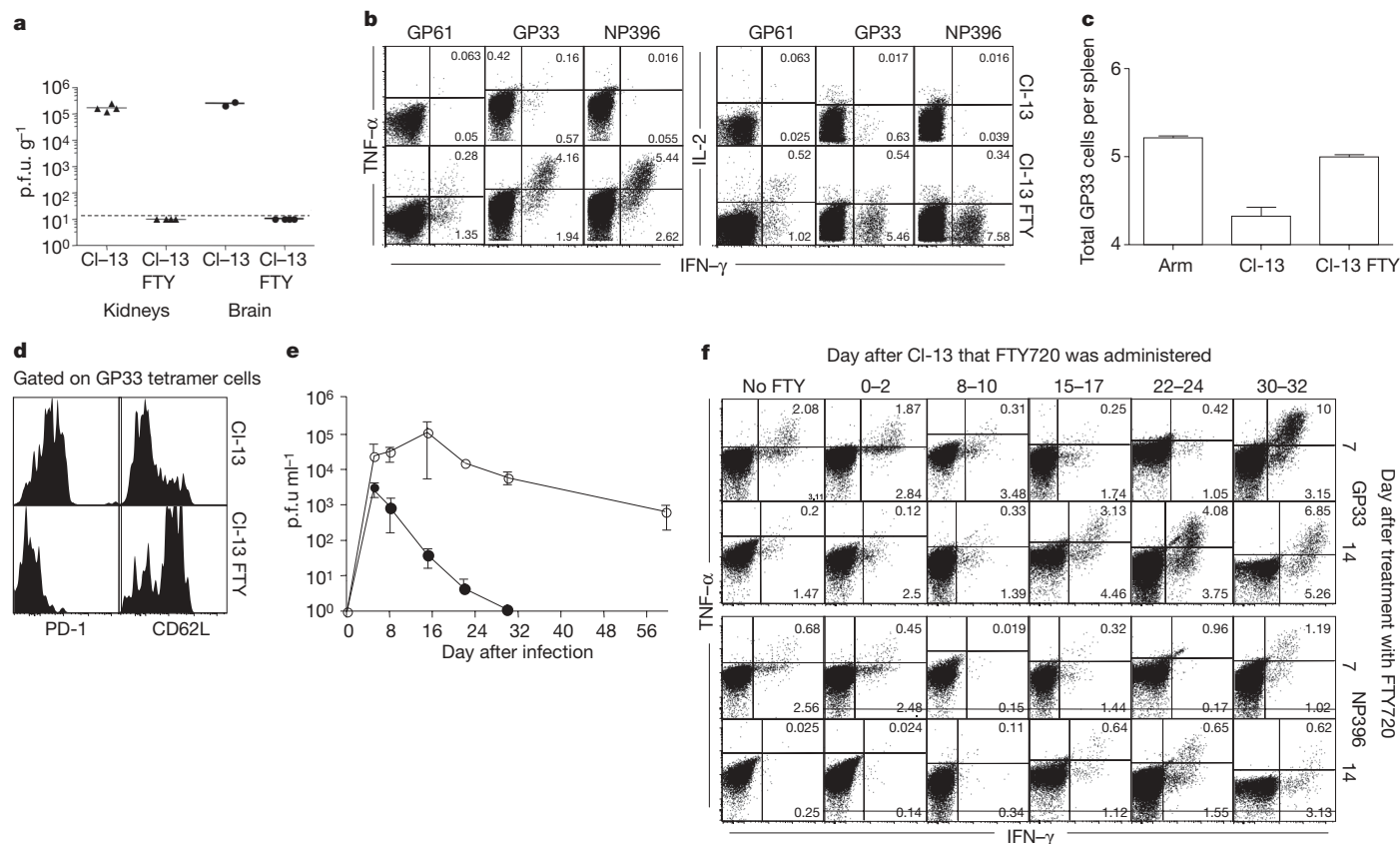
$P = 0.0256$  day 15,  $P = 0.0010$  day 22 and  $P = 0.0058$  day 30. **c–e**, T-cell responses to LCMV epitopes were analysed by intracellular cytokine staining in mice infected with Armstrong (Arm), clone 13 (CI-13) and FTY720-treated clone 13 (CI-13 FTY). CD8 responses to GP33 (**c**), to NP396 (**d**), and CD4 responses to GP61 (**e**) (day 30 only) are shown. **f**, GP33 and NP396 CD8 T-cell responses were analysed by intracellular cytokine staining for production of IL-2 and IFN- $\gamma$  at 60 days after infection. For all panels, mice were used at 3–5 per group; data are representative of 2–4 experiments. The numbers on the fluorescence-activated cell sorting (FACS) plot indicate the percentage of total CD8 or CD4 T cells in each quadrant. Error bars are s.d.

thymectomized before infection (Supplementary Fig. 7). In the thymectomized mice, the kinetics of expansion of LCMV NP396-specific cells was slower than cells specific for GP33. These results are consistent with the interpretation that at least a fraction of the functional cell populations seen after FTY720 treatment derives from previously dysfunctional LCMV-specific cells.

Because CD4 T cells are important for the maintenance of CD8 T-cell responses, we tested whether FTY720 can enhance immunity in mice depleted of CD4 T cells at the time of infection. CD4 T-cell depletion ablated the immune enhancing effects of FTY720 treatment and there was no significant reduction of viral titres (Fig. 3e), nor were CD8 T-cell responses enhanced when CD4 T cells were not present (Fig. 3f). However, CD4 T cells return to the circulation between days 10 to 15 after antibody-mediated depletion, and on their return, treatment with FTY720 (beginning on day 15 or later) restored LCMV-specific T-cell responses (Fig. 3f). CD4-depleted but not untreated control mice showed a progressive loss of cells producing IFN- $\gamma$  in response to stimulation with GP33 and NP396

peptides (data not shown), in agreement with previously published data<sup>16</sup>. These data show that CD4 T cells are necessary for the immune enhancing effects of FTY720 and demonstrate that FTY720-induced clearance of clone 13 is mediated by the immune response and is not directly due to any antiviral effects of the drug.

Several other approaches have been used to treat chronic LCMV infections, including using anti-PD-L1 (ref. 17), anti-IL-10R (ref. 18) or anti-CD70 (ref. 19) antibodies, and therapeutic vaccination combined with antibody blockade<sup>20,21</sup>, but none was as successful at clearing an established chronic infection as the transient use of FTY720. Although we were motivated to test FTY720 on the basis of our observations of reduced lymphopenia during clone 13 infection, we do not yet know the exact mechanism by which it is promoting clearance of chronic LCMV. On the basis of the known targets of FTY720 (ref. 10), we have no reason to suspect it is mechanistically connected to any of the previously described therapies. In this study we have shown that treatment with FTY720 leads to the restoration of functional antigen-specific CD8 and CD4 T cells. The restoration of



**Figure 3 | Low-dose short-term FTY720 treatment clears an established persistent clone 13 infection.** Wild-type mice were infected with LCMV clone 13 at  $2 \times 10^6$  p.f.u. intravenously and treated with vehicle or FTY720 at  $4 \mu\text{g kg}^{-1}$  intravenously on days 30, 31 and 32 after infection. **a**, Viral titres in the kidneys and brain were measured using a standard plaque assay at day 30 after FTY720 treatment (60 days after infection;  $P = 0.0008$ , unpaired *t*-test). **b**, IFN- $\gamma$ , IL-2 and TNF- $\alpha$  production by LCMV-specific T cells in clone-13-infected (CI-13) and FTY720-treated clone-13-infected (CI-13 FTY) mice were analysed by intracellular cytokine staining. **c**, Total GP33-tetramer-positive CD8 T cells in the spleen 4 weeks after FTY720 treatment. **d**, PD-1 and CD62L expression on GP33-tetramer-positive cells in the lymph nodes at day 30 after FTY720 treatment (day 60 after infection). **e**, Serum viral titres in clone-13-infected mice that were either treated with FTY720 (filled

circles), or CD4 depleted and FTY720 treated (open circles) at the time of infection. CD4 depletion was performed by administration of  $500 \mu\text{g}$  of GK1.5 monoclonal antibody at days 0 and 1; FTY720 was administered intravenously at  $4 \mu\text{g kg}^{-1}$  on days 0, 1 and 2 after infection. **f**, GP33 and NP396 T-cell responses—measured by IFN- $\gamma$  and TNF- $\alpha$  production—were analysed in the spleens of clone-13-infected mice that were depleted of CD4 T cells at the time of infection and treated with FTY720 at various time points after infection as indicated; T-cell responses were analysed 7 and 14 days after FTY720 treatment. GP33 and NP396 CD8 T cells from LCMV-infected mice were stimulated with peptide and analysed for cytokine production 1 and 2 weeks after FTY720 treatment. All data are representative of 2–3 experiments;  $n = 3$ –5 per group. Error bars are s.d.

CD8 T cells is dependent on CD4 T cells and in thymectomized mice there is a delay in restoration. Therefore, it is possible that a *de novo* CD4 T-cell response is necessary to restore dysfunctional CD8 T cells. In addition, this treatment was not effective in B-cell knockout mice (data not shown), although it is not clear whether this is due to the lack of B cells or the lack of normal immune architecture<sup>22</sup>.

It remains unclear as to how a drug well known for its immunosuppressive effects is also able to induce immune-mediated clearance of a chronic viral infection. In the setting of multiple sclerosis and prevention of transplant rejection, FTY720 is thought to act by keeping cells sequestered in the lymph nodes, away from the peripheral sites of pathology. In contrast, during an infection lymph nodes are a site of immune induction, and in the context of an LCMV infection, lymph nodes are also a significant site of virus production and pathology. Because our current data on FTY720-induced clearance of a virus are limited to the single example of the LCMV model, it remains to be determined whether the effect translates to the successful treatment of other chronic infections, including those that afflict humans.

## METHODS SUMMARY

**Mice and infections.** C57BL/6, C57BL/6-thymectomized mice and B6.PL-*Thy1<sup>0</sup>/CyJ* mice were purchased from the Jackson Laboratory; F5 mice were provided by D. Moskophidis; P14 mice were a gift from the Ahmed laboratory.

LCMV Armstrong and clone 13 strains were provided by the Ahmed laboratory and administered intravenously at  $2 \times 10^6$  plaque-forming units (p.f.u.). Where indicated, CD4 T cells were depleted as described<sup>16</sup>.

**Administration of FTY720.** FTY720 (Cayman Chemical) was reconstituted in DMSO and diluted with water and was administered intravenously at various doses for 3 days in a total volume of  $200 \mu\text{l}$ . Control animals received mock treated water.

**Cell enumeration and staining.** Cells were counted using a standard haemocytometer. Flow cytometric analysis was performed on a LSRII (Becton Dickinson). All monoclonal antibodies used were from Becton Dickinson, CalTag or eBioscience. Tetramers were produced as described<sup>23</sup>.

**Type I interferon measurement.** A type I interferon bioassay was used to measure IFN-I as described<sup>24</sup>. IFN- $\alpha$  (Biomedical Laboratories PBL) was used as a control.

**SIP measurement: mass spectrometry.** Serum and lymph nodes were collected from three mice and were pooled. Lymph nodes were weighed and homogenized. Sphingolipid extraction was performed using methanol and chloroform followed by potassium hydroxide. C17-sphingosine-1-phosphate (Avanti Polar Lipids) was used as an internal control. After extraction, sphingoid base 1-phosphate was quantified using liquid chromatography tandem mass spectrometry on reverse phase (C18) chromatography with a Perkin Elmer Series 200 HPLC system that interfaces with an API 2000 (Applied Biosystems) triple quadrupole mass spectrometer<sup>25</sup>.

**Intracellular cytokine assay.** Intracellular cytokine staining was performed as described<sup>7</sup>.

**Statistics.** A two-tailed Student's *t*-test determined all probabilities, calculated by inStat.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 31 March; accepted 26 June 2008.**

- Ahmed, R., Byrne, J. A. & Oldstone, M. B. Virus specificity of cytotoxic T lymphocytes generated during acute lymphocytic choriomeningitis virus infection: role of the H-2 region in determining cross-reactivity for different lymphocytic choriomeningitis virus strains. *J. Virol.* **51**, 34–41 (1984).
- Matloubian, M., Concepcion, R. J. & Ahmed, R. CD4<sup>+</sup> T cells are required to sustain CD8<sup>+</sup> cytotoxic T-cell responses during chronic viral infection. *J. Virol.* **68**, 8056–8063 (1994).
- Kunzendorf, U., Ziegler, E. & Kabelitz, D. FTY720—the first compound of a new promising class of immunosuppressive drugs. *Nephrol. Dial. Transplant.* **19**, 1677–1681 (2004).
- Chisari, F. V. & Ferrari, C. Hepatitis B virus immunopathology. *Springer Semin. Immunopathol.* **17**, 261–281 (1995).
- Day, C. L. *et al.* PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. *Nature* **443**, 350–354 (2006).
- Folgori, A. *et al.* Early impairment of hepatitis C virus specific T cell proliferation during acute infection leads to failure of viral clearance. *Gut* **55**, 1012–1019 (2006).
- Wherry, E. J., Blattman, J. N., Murali-Krishna, K., van der Most, R. & Ahmed, R. Viral persistence alters CD8 T-cell immunodominance and tissue distribution and results in distinct stages of functional impairment. *J. Virol.* **77**, 4911–4927 (2003).
- Shiow, L. R. *et al.* CD69 acts downstream of interferon- $\alpha/\beta$  to inhibit S1P<sub>1</sub> and lymphocyte egress from lymphoid organs. *Nature* **440**, 540–544 (2006).
- Mamalaki, C. *et al.* Positive and negative selection in transgenic mice expressing a T-cell receptor specific for influenza nucleoprotein and endogenous superantigen. *Dev. Immunol.* **3**, 159–174 (1993).
- Brinkmann, V. *et al.* The immune modulator FTY720 targets sphingosine 1-phosphate receptors. *J. Biol. Chem.* **277**, 21453–21457 (2002).
- Chiba, K. FTY720, a new class of immunomodulator, inhibits lymphocyte egress from secondary lymphoid tissues and thymus by agonistic activity at sphingosine 1-phosphate receptors. *Pharmacol. Ther.* **108**, 308–319 (2005).
- Kappos, L. *et al.* Oral fingolimod (FTY720) for relapsing multiple sclerosis. *N. Engl. J. Med.* **355**, 1124–1140 (2006).
- Budde, K. *et al.* First human trial of FTY720, a novel immunomodulator, in stable renal transplant patients. *J. Am. Soc. Nephrol.* **13**, 1073–1083 (2002).
- Pinschewer, D. D. *et al.* FTY720 immunosuppression impairs effector T cell peripheral homing without affecting induction, expansion, and memory. *J. Immunol.* **164**, 5761–5770 (2000).
- Moskophidis, D., Lechner, F., Pircher, H. & Zinkernagel, R. M. Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. *Nature* **362**, 758–761 (1993).
- Zajac, A. J. *et al.* Viral immune evasion due to persistence of activated T cells without effector function. *J. Exp. Med.* **188**, 2205–2213 (1998).
- Barber, D. L. *et al.* Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature* **439**, 682–687 (2006).
- Brooks, D. G. *et al.* Interleukin-10 determines viral clearance or persistence *in vivo*. *Nature Med.* **12**, 1301–1309 (2006).
- Matter, M., Odermatt, B., Yagita, H., Nuoffer, J. M. & Ochsenbein, A. F. Elimination of chronic viral infection by blocking CD27 signaling. *J. Exp. Med.* **203**, 2145–2155 (2006).
- Brooks, D. G., Lee, A. M., Elsaesser, H., McGavern, D. B. & Oldstone, M. B. IL-10 blockade facilitates DNA vaccine-induced T cell responses and enhances clearance of persistent virus infection. *J. Exp. Med.* **205**, 533–541 (2008).
- Ha, S. J. *et al.* Enhancing therapeutic vaccination by blocking PD-1-mediated inhibitory signals during chronic infection. *J. Exp. Med.* **205**, 543–555 (2008).
- Muller, S. *et al.* Role of an intact splenic microarchitecture in early lymphocytic choriomeningitis virus production. *J. Virol.* **76**, 2375–2383 (2002).
- Altman, J. D. *et al.* Phenotypic analysis of antigen-specific T lymphocytes. *Science* **274**, 94–96 (1996).
- Cembrzynska-Nowak, M. Different antiviral activity and cell specificity of interferon preparations produced by mouse peritoneal cells at 37 degrees C and at 26 degrees C. *Arch. Immunol. Ther. Exp. (Warsz.)* **37**, 499–502 (1989).
- Merrill, A. H. Jr, Sullards, M. C., Allegood, J. C., Kelly, S. & Wang, E. Sphingolipidomics: high-throughput, structure-specific, and quantitative analysis of sphingolipids by liquid chromatography tandem mass spectrometry. *Methods* **36**, 207–224 (2005).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We would like to thank R. Ahmed for his support and guidance during the project; D. Liotta for the use of his mass spectrometer; and the members of the Ahmed and the Altman laboratories for discussions. This work was supported by National Institute of Health grant numbers AI042373 (to J.D.A.) and 5F32AI062002 (M.P.-L.).

**Author Contributions** M.P.-L. originated and designed the project, analysed the results and co-wrote the manuscript with J.D.A. who also provided guidance and support throughout the project. N.B.M. performed viral load assays. S.T.P. measured S1P levels. P.A.R. was involved in preliminary experiments.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to J.D.A. ([jaltman@rmy.emory.edu](mailto:jaltman@rmy.emory.edu)) or M.P.-L. ([mflan@emory.edu](mailto:mflan@emory.edu)).

## METHODS

**Mice, virus, infections and treatments.** C57BL/6 (Ly5.2) and thymectomized mice were purchased from the Jackson Laboratory. P14 mice were a gift from the Ahmed laboratory and were housed in our animal facility. All infections were performed intravenously at  $2 \times 10^6$  p.f.u. of LCMV clone 13 or LCMV Armstrong. Where indicated, CD4 T cells were depleted by injecting 500  $\mu$ g GK1.5 in PBS on the day of infection and 1 day after infection. Viral titres in the serum and tissue homogenates were performed on Vero cells as described<sup>1</sup>. FTY720 (Cayman Chemical) was dissolved at 20 mg ml<sup>-1</sup> in DMSO. FTY720 was diluted in water and administered in a total volume of 200  $\mu$ l, in a range of 800 ng kg<sup>-1</sup> to 1 mg kg<sup>-1</sup>, intravenously.

**Tissue preparation and cell enumeration.** Blood was collected either with or without heparin tubes and spun down to collect serum. Serum was used to measure type I IFNs, S1P levels and viral titres. Peripheral blood mononuclear cells were separated from heparin blood using Lympholyte (Cedarlane Laboratories Limited, CL5115). Tissues were collected and lymphocytes were separated by force; when indicated collagenase type 4 (Worthington 46E8824) was used at 1 mg ml<sup>-1</sup>. Cells were counted visually using a haemocytometer.

**Flow cytometry and *in vitro* stimulations.** Anti-CD8, anti-PD-1, anti-CD69, anti-IFN- $\gamma$ , anti-TNF- $\alpha$  and IL-2 were purchased from BD Biosciences. Anti-CD4, anti-CD19 and anti-thy antibodies were purchased from eBioscience. MHC class I tetramers were manufactured as described<sup>24</sup>. To quantify LCMV specific CD8 T cells, DbNP396–404, DbGP33–41 and DbGP276–286 tetramers were used. Intracellular cytokine staining was performed according to the manufacturer's protocol following a 6 h stimulation of cells with 0.2  $\mu$ g ml<sup>-1</sup> of the indicated peptide.

**Type I IFN bioassay.** IFN-I bioactivity was measured using the L929 bioassay<sup>24</sup>. L929 cells ( $3 \times 10^4$  L929 cells per 100  $\mu$ l) in MEM with 5% FCS were added to a 96-well plate and incubated overnight at 37 °C in successive twofold dilutions of serum samples. The murine IFN standard (Biomedical Laboratories PBL, 12100-1) was used at a final concentration of 100 U ml<sup>-1</sup>. Twenty-four hours later, vesicular stomatitis virus was added to the wells at 50 p.f.u. per well and the cells were incubated for 3 days. The dilution mediating 50% protection was defined as 1 U of IFN-I per ml.

**S1P measurement: mass spectrometry.** Serum and lymph nodes from three mice were collected and pooled. Lymph nodes were weighed and homogenized. Sphingolipid extraction was performed using methanol and chloroform followed by potassium hydroxide. C17-sphingosine-1-phosphate (Avanti Polar Lipids) served as an internal control. Following extraction, sphingoid base 1-phosphate was quantified using liquid chromatography tandem mass spectrometry on reverse phase (C18) chromatography with a Perkin Elmer Series 200 HPLC system that interfaces with an API 2000 (Applied Biosystems) triple quadrupole mass spectrometer<sup>25</sup>.

# A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA

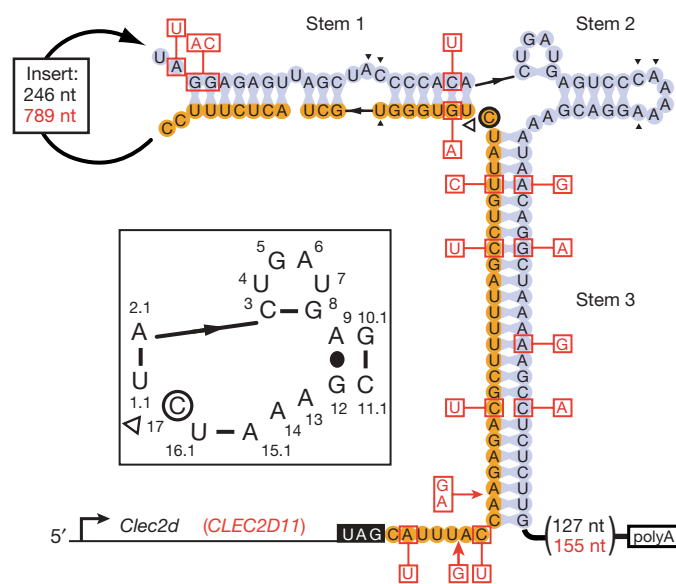
Monika Martick<sup>1\*</sup>, Lucas H. Horan<sup>1\*</sup>, Harry F. Noller<sup>1</sup> & William G. Scott<sup>1</sup>

Structured RNAs embedded in the untranslated regions (UTRs) of messenger RNAs can regulate gene expression. In bacteria, control of a metabolite gene is mediated by the self-cleaving activity of a ribozyme embedded in its 5' UTR<sup>1</sup>. This discovery has raised the question of whether gene-regulating ribozymes also exist in eukaryotic mRNAs. Here we show that highly active hammerhead ribozymes<sup>2,3</sup> are present in the 3' UTRs of rodent C-type lectin type II (*Clec2*) genes<sup>4–7</sup>. Using a hammerhead RNA motif search with relaxed delimitation of the non-conserved regions, we detected ribozyme sequences in which the invariant regions, in contrast to the previously identified continuous hammerheads<sup>8–10</sup>, occur as two fragments separated by hundreds of nucleotides. Notably, a fragment pair can assemble to form an active hammerhead ribozyme structure between the translation termination and the polyadenylation signals within the 3' UTR. We demonstrate that this hammerhead structure can self-cleave both *in vitro* and *in vivo*, and is able to reduce protein expression in mouse cells. These results indicate that an unrecognized mechanism of post-transcriptional gene regulation involving association of discontinuous ribozyme sequences within an mRNA may be modulating the expression of several CLEC2 proteins that function in bone remodelling and the immune response of several mammals.

The hammerhead ribozyme is a small, self-cleaving motif composed of a three-helical junction with a core of invariant nucleotides required for activity. To identify hammerhead ribozymes in mammalian mRNAs, we searched mRNA sequence databases using a pattern descriptor that allowed for insertions of up to 5,000 nucleotides at the ends of stem 1 or stem 3 (Supplementary Fig. 1)<sup>11,12</sup>. Three hammerhead ribozymes were identified in the 3' UTRs of known rodent mRNAs. Two are found embedded in the transcripts of mouse *Clec2d* (osteoclast inhibitory lectin, also known as *Ocil*, *Clr-b* and *Clec2d8*)<sup>13</sup> and its paralogue *Clec2e* (also known as *Clra* and *Clec2d7*)<sup>14</sup>, genes that belong to a group of phylogenetically related sequences within the natural killer receptor gene complex of chromosome 6. The third ribozyme is found in rat *CLEC2D11* (ref. 7)—a homologue of mouse *Clec2d*—which resides in the syntenic natural killer receptor gene complex region on chromosome 4. We extended our search to the genomic sequences of other organisms using the UCSC genome browser's comparative genomics tool<sup>15</sup>. Alignments using the natural killer receptor gene complex regions of mouse and rat led to the identification of nine candidate hammerhead ribozymes: four in the 3' regions of predicted rat, horse and platypus *Clec2*-like genes, and five in the unannotated regions of five other mammalian genomes (Supplementary Fig. 2).

Unlike most known self-cleaving RNA motifs that are contiguous<sup>8–10,16–19</sup>, the hammerhead ribozymes identified here (referred to as CLEC2 ribozymes) are split into two fragments separated by a long ribozyme-unrelated insertion in the stem-1-capping

loop. The insertion (which is 250 and 696 nucleotides in mouse *Clec2d* and *Clec2e*, respectively, and 145–1,739 nucleotides in other candidate ribozyme sequences) segregates the upstream substrate region, residing within 7–44 nucleotides of the stop codon, from the downstream enzyme fragment (Fig. 1 and Supplementary Fig. 2). When assembled from substrate and enzyme fragments, the secondary structure reflects a characteristic hammerhead ribozyme core of fifteen conserved nucleotides flanked by three helices. Comparison of all twelve CLEC2 ribozymes showed conservation of stem 2 as well as the presence of compensatory mutations maintaining stem 1 and the atypically long stem 3. Stem 1 and stem 2 contain secondary structure elements required to form a tertiary contact known to enhance catalysis greatly (Fig. 1 and Supplementary Fig. 2)<sup>3,20,21</sup>. Six loop and bulge residues (Fig. 1)—necessary for the active structure



**Figure 1 | Sequence arrangement and secondary structure model of the rodent *Clec2d* hammerhead ribozymes.** The mouse ribozyme sequence is shown in black and the rat ribozyme sequence length, single nucleotide, and base pair differences are denoted in red. The stop codon is shown in white. The substrate and enzyme sequences are shown on orange and blue backgrounds, respectively. The insertion sequence separating two ribozyme parts is abridged with a thick arrow. The predicted cleavage site (white arrowhead) is 3' of the active site cytosine (circled). The three-helical junction, composed of conserved (with the exception of 2.1 and 1.1) nucleotides (nt) that make catalytically important interactions, is shown in greater detail (together with canonical numbering scheme) in the inset. The small black arrowheads indicate conserved nucleotides of the catalytically important loop/bulge interactions.

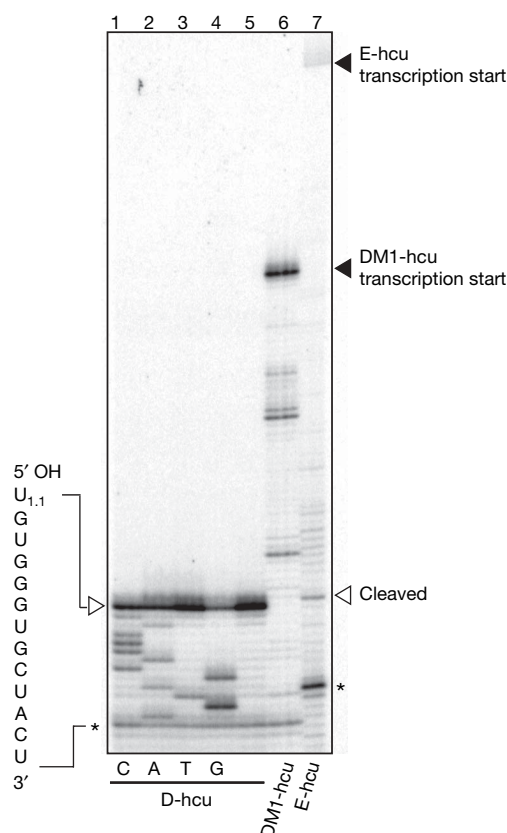
<sup>1</sup>Center for Molecular Biology of RNA, University of California, Santa Cruz, California 95064, USA.

\*These authors contributed equally to this work.

stabilization and optimal catalysis in the *Schistosoma mansoni* hammerhead ribozyme<sup>22</sup>—are also found in most CLEC2 ribozyme sequences. Taken together, the secondary structure features suggest that the 3' UTR-embedded hammerhead ribozyme motifs form catalytically active structures despite their discontinuity.

To test whether the hammerhead ribozymes embedded within the 3' UTRs possess catalytic activity, we examined the *in vitro* transcription products of mouse 3' UTRs from *Clec2d* (D-hcu, hammerhead containing UTR) and *Clec2e* (E-hcu). Cleavage of both 3' UTR RNAs occurs during the course of the transcription reaction (data not shown) and at the predicted hammerhead ribozyme cleavage sites (Fig. 2). The DM1-hcu construct—which is expected to create an inactive hammerhead ribozyme due to a triple mutation that prevents the 'substrate' from being cleaved—yields an intact 3' UTR transcript, thus confirming that the cleavage is due to the embedded hammerhead ribozyme activity. The mouse *Clec2e*-derived 3' UTR cleaves to a lesser extent than D-hcu, perhaps because of misfolding effects due to a ~450 nucleotide longer intervening sequence and/or weakened secondary structure in the stem 1 and stem 3 regions (Supplementary Fig. 3). These results demonstrate that the discontinuous 'substrate' and 'enzyme' fragments form a fully active hammerhead ribozyme.

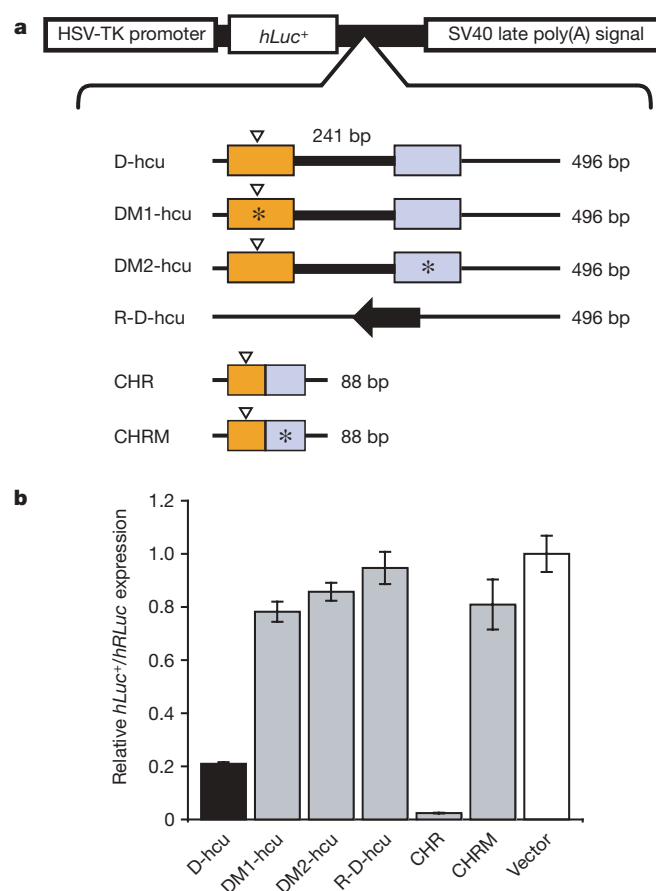
To determine whether the D-hcu sequence can reduce gene expression by forming an active ribozyme *in vivo*, it was incorporated into a dual-luciferase gene expression vector and assayed for activity in NIH 3T3 cells (Fig. 3). D-hcu reduces luciferase expression by 80% when placed downstream from the luciferase reporter (Fig. 3b). Nonspecific downregulation effects were ruled out because inversion of the same sequence (R-D-hcu) restores luciferase expression levels



**Figure 2 | The discontinuous hammerhead ribozyme-containing 3' UTRs (hcu) self-cleave *in vitro*.** Reverse transcriptase primer extensions of *in vitro* transcribed mouse *Clec2d* (D-hcu, lane 5) and *Clec2e* (E-hcu, lane 7) 3' UTRs show that both RNAs are cleaved at a single location (open arrowheads). D-hcu cleavage site 5' of U<sub>1.1</sub> was identified by dideoxy sequencing of transcription products. Uncleaved E-hcu and triple mutant DM1-hcu are indicated with a black arrowhead. Asterisks indicate non-extended labelled primers.

to that of the positive control. To verify that the reduction of gene expression was caused by the embedded ribozyme, mutations specifically targeting hammerhead ribozyme activity were introduced into D-hcu. Both a single transversion mutation at position G8 in the enzyme fragment (DM2-hcu), previously shown to compromise hammerhead ribozyme activity<sup>3</sup>, and a triple mutation in the substrate fragment (DM1-hcu) restore gene expression to control levels, indicating that the hammerhead ribozyme embedded in the mouse *Clec2d* 3' UTR is responsible for reducing reporter protein expression in mouse cells.

RNA from cells transfected with either wild-type (D-hcu) or mutant (DM1-hcu) dual luciferase reporter vectors was compared using reverse transcription–polymerase chain reaction (RT–PCR) to assess the effect of the discontinuous ribozyme-containing 3' UTR on mRNA integrity. A mixture of two primers, each recognizing one of two reporter genes transcribed from the same vector, was used for first-strand complementary DNA synthesis (Fig. 4a). Amplification of the region spanning the ribozyme cleavage site indicates that



**Figure 3 | The discontinuous mouse *Clec2* hammerhead ribozymes embedded in the 3' UTRs downregulate protein expression *in vivo*.** **a**, Layout of the constructs (top) and the schematic of hammerhead ribozyme-containing 3' UTR (hcu) sequences used for expression. The colours of substrate (orange) and enzyme (blue) regions correspond to those of the structure in Fig. 1 and the insertion is highlighted by a thick line. The cleavage sites are indicated by white arrowheads and asterisks denote the positions of mutations. CHR (control hammerhead ribozyme) and CHRM (G8 to C8 mutant of CHR) are the controls; bp, base pairs. **b**, *In vivo* analysis of firefly luciferase protein expression from different 'hcu' constructs. Black denotes expression from the vector containing wild type 3' UTR; grey represents mutants and controls; and white indicates a vector without an insert as a negative control. Relative firefly luciferase (*hLuc*<sup>+</sup>) expression was determined using a dual luciferase assay using the *Renilla* luciferase (*hRLuc*) expression for normalization. The results shown are means  $\pm$  s.e.m. of triplicate experiments; the error bars for the vector value are averages of s.e.m. from three assays.

mouse *Clec2d* 3' UTR self-cleaves via the hammerhead ribozyme activity (Fig. 4b). The uncleaved product of D-hcu is at least tenfold reduced from that of DM1-hcu, in accordance with results of the protein expression assay (Fig. 4c). Previous studies have shown that ribozymes artificially engineered into 3' UTRs can downregulate gene expression<sup>23</sup> by promoting rapid destruction of the transcript by cytoplasmic RNA degradation machinery<sup>24</sup>. Consistent with these observations, the catalytic activity of mouse *Clec2d* 3' UTR correlates with the overall reduction of firefly luciferase transcript levels, as compared to those of the mutant UTR-containing mRNA, indicating that the cleaved *Clec2* transcript is degraded *in vivo*.

Structural conservation observed throughout the identified sequences, in conjunction with the tight association of the ribozyme with *Clec2*-like genes, suggests that the discontinuous ribozymes are orthologous to the mouse hammerhead motif characterized here. Within the rodent lineage, the rat *CLEC2D11* ribozyme sequence shares 86% homology with the mouse *Clec2d* ribozyme. Similarly, two predicted rat *Clec2d* paralogues are closely (61% and 67%) related to the mouse *Clec2e* ribozyme (Supplementary Fig. 2). These sequence relationships positively correlate with phylogenetic evidence for multiple duplication events of rodent *Clec2* genes<sup>7</sup>. Further genome searches using separate parts of the hammerhead ribozyme detected four additional mouse *Clec2* paralogues that contain substrate-like, but not enzyme, sequences in their 3' UTRs (data not shown). These partial motifs are also positioned proximally to

the stop codon and share extensive primary sequence identity with the substrate regions of functional ribozymes of either mouse *Clec2d* or *Clec2e*. However, it is not clear if the incomplete ribozyme sequences are functional substrates for the *Clec2d*- and *Clec2e*-derived trans-acting ribozyme elements, or are non-functional evolutionary relics of the rodent *Clec2* gene expansion.

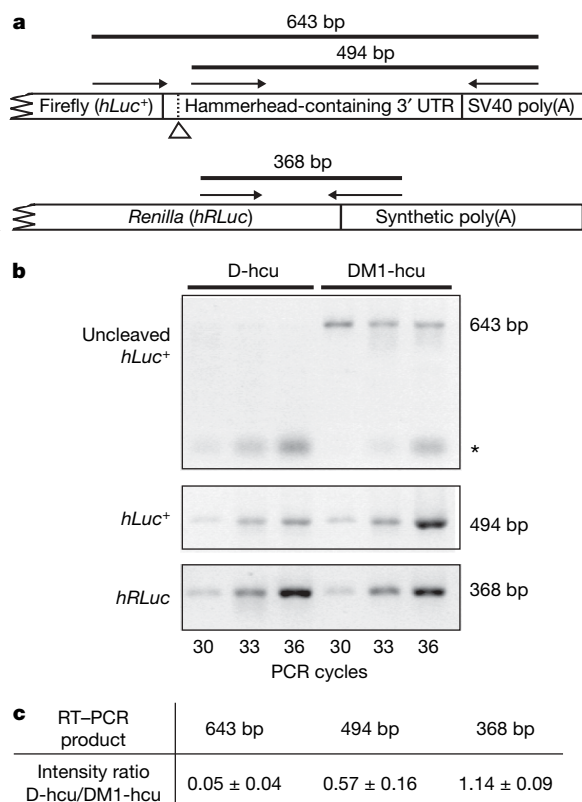
Identification of CLEC2 ribozymes in the genomes of mammalian species other than rodents (notably platypus) indicates that these ribozymes share an early mammalian ancestor. Active hammerhead motifs have been found in several divergent species, such as newts<sup>10</sup>, plants<sup>8</sup> and viroids<sup>10</sup>. However, only the *Schistosoma* satellite DNA is known to encode hammerhead ribozyme sequences that have significant homology to the CLEC2 ribozymes<sup>3,18</sup>. Although schistosomes parasitize rodents, their ribozymes are found in lineages more modern than the time of divergence of mouse and rat species<sup>18,25</sup>, suggesting a horizontal gene transfer of the hammerhead sequence from host to the parasite genome as the probable scenario.

In eukaryotes, post-transcriptional regulation of gene expression through the 3' UTR involves an interaction between mRNA-stabilizing and -destabilizing protein-based processes<sup>26</sup>. Together with evidence of cross-species secondary structure conservation, our results indicate that the CLEC2 hammerhead ribozymes have roles analogous to those of destabilizing protein factors. Paired with a yet undiscovered antagonistic process, the hammerhead motif may constitute a system of rapid regulation for CLEC2d, a transmembrane C-type lectin that has a dual role: the inhibition of natural killer (NK)-cell-mediated lysis through the interaction with the NK cell receptor NKR-P1, and the inhibition of osteoclast formation<sup>4,6</sup>.

The core sequences of CLEC2 ribozymes resemble those of the smaller members of its ribozyme class, but the discontinuous arrangement of the substrate and enzyme regions exhibit properties previously associated only with catalytic introns. Although functioning through a fundamentally different mechanism and residing in transcripts distinct from processed mRNA, catalytic sequences of some self-splicing group I and II introns span long stretches of RNA<sup>19</sup> on a scale analogous to that of the ribozymes described here. A question remains as to whether CLEC2 ribozymes can also process substrates encoded on detached transcripts similarly to RNase P, which possesses trans-cleaving properties unique among naturally occurring ribozymes<sup>19</sup>. Most self-cleaving RNAs, however, are considered to be compact sequences suitable for single-step functions such as multimer-to-monomer conversion of genomes in the course of viroid and satellite RNA replication, or for response to small molecule binding such as in metabolite-dependent gene regulation in bacteria. However, in the context of complex gene regulatory systems such as those of mammalian cells, the discontinuous configuration of the CLEC2 ribozyme sequences may provide an opportunity for allosteric regulation.

## METHODS SUMMARY

mRNA sequence data from the Ensembl database (release 48, <http://www.ensembl.org>) were searched using the RNABOB program<sup>11,12</sup> with a descriptor (Supplementary Fig. 1) for the hammerhead ribozyme motif. Orthologues of the rodent CLEC2 ribozymes were identified using BLAT and the UCSC Genome Browser (<http://genome.ucsc.edu/>)<sup>15,27</sup>. For *in vitro* RNA synthesis and analysis, the *Clec2* 3' UTR-containing plasmids were linearized and transcribed using T7 RNA polymerase at a pH of 8.0 and in the presence of 28 mM MgCl<sub>2</sub>. Reverse transcriptase primer extensions and RNA sequencing were adapted for a [<sup>32</sup>P]-end-labelled primer from previously published methods<sup>28</sup>. For the expression assays, the NIH 3T3 mouse fibroblast cells were transiently transfected with *Clec2* 3' UTR-containing psiCHECK-2 reporter plasmids (Promega) and analysed using the Dual-Glo Luciferase Assay System (Promega). Firefly reporter expression was normalized using expression of the *Renilla* luciferase encoded on the same plasmid. RNA from the transfected cells was purified using Trizol (Sigma). For RT-PCR, cDNA was synthesized using primers specific to plasmid-derived transcripts using previous methods<sup>28</sup>. PCR was performed with Accuprime Taq (Invitrogen). PCR products were resolved on an agarose gel and the product quantification was performed using ImageQuant 5.2 (Molecular Dynamics).



**Figure 4 | RT-PCR analysis of the *in vivo* expression products.** **a**, Schematic of the relative positions of the RT-PCR primers. The hammerhead-specific cleavage site is indicated by an open arrowhead. The 368 base pair (bp) product acts as the internal control for transfection efficiency and the extent of transcription. **b**, RT-PCR of RNA isolated from NIH 3T3 cells that have been transfected with pD-hcu or pDM1-hcu. The asterisk indicates nonspecific PCR products. **c**, Quantification of the difference in PCR product intensity between D-hcu and DM1-hcu. The product intensity was used to determine the ratio of D-hcu/DM1-hcu at each listed cycle. Each value represents the mean ± s.d. of intensity ratios from three listed cycles.

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Received 10 January; accepted 30 May 2008.**

**Published online 9 July 2008.**

- Winkler, W. C. *et al.* Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281–286 (2004).
- Khvorova, A., Lescoute, A., Westhof, E. & Jayasena, S. D. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nature Struct. Biol.* **10**, 708–712 (2003).
- Martick, M. & Scott, W. G. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **126**, 309–320 (2006).
- Zhou, H. *et al.* Osteoclast inhibitory lectin, a family of new osteoclast inhibitors. *J. Biol. Chem.* **277**, 48808–48815 (2002).
- Plougastel, B., Dubbelde, C. & Yokoyama, W. M. Cloning of *Clr*, a new family of lectin-like genes localized between mouse *Nkrp1a* and *Cd69*. *Immunogenetics* **53**, 209–214 (2001).
- Carlyle, J. R. *et al.* Missing self-recognition of Ocil/Clr-b by inhibitory NKR-P1 natural killer cell receptors. *Proc. Natl Acad. Sci. USA* **101**, 3527–3532 (2004).
- Hao, L., Klein, J. & Nei, M. Heterogeneous but conserved natural killer receptor gene complexes in four major orders of mammals. *Proc. Natl Acad. Sci. USA* **103**, 3192–3197 (2006).
- Przybilski, R. *et al.* Functional hammerhead ribozymes naturally encoded in the genome of *Arabidopsis thaliana*. *Plant Cell* **17**, 1877–1885 (2005).
- Rojas, A. A. *et al.* Hammerhead-mediated processing of satellite pDo500 family transcripts from Dolichopoda cave crickets. *Nucleic Acids Res.* **28**, 4037–4043 (2000).
- Flores, R. *et al.* Hammerhead ribozyme structure and function in plant RNA replication. *Methods Enzymol.* **341**, 540–552 (2001).
- Gautheret, D., Major, F. & Cedergren, R. Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA. *Comput. Appl. Biosci.* **6**, 325–331 (1990).
- Eddy, S. R. RNABOB: a program to search for RNA secondary structure motifs in sequence databases. (<http://selab.janelia.org/software.html>) (2008).
- Zhou, H. *et al.* A novel osteoblast-derived C-type lectin that inhibits osteoclast formation. *J. Biol. Chem.* **276**, 14916–14923 (2001).
- Carlyle, J. R. *et al.* Molecular and genetic basis for strain-dependent NK1.1 alloreactivity of mouse NK cells. *J. Immunol.* **176**, 7511–7524 (2006).
- Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
- Salehi-Ashtiani, K., Luptak, A., Litovchick, A. & Szostak, J. W. A genomewide search for ribozymes reveals an HDV-like sequence in the human *CPEB3* gene. *Science* **313**, 1788–1792 (2006).
- Teixeira, A. *et al.* Autocatalytic RNA cleavage in the human  $\beta$ -globin pre-mRNA promotes transcription termination. *Nature* **432**, 526–530 (2004).
- Ferbeyre, G., Smith, J. M. & Cedergren, R. Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol. Cell. Biol.* **18**, 3880–3888 (1998).
- Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nature Rev. Genet.* **8**, 776–790 (2007).
- Canny, M. D. *et al.* Fast cleavage kinetics of a natural hammerhead ribozyme. *J. Am. Chem. Soc.* **126**, 10848–10849 (2004).
- Osborne, E. M., Schaak, J. E. & Deroose, V. J. Characterization of a native hammerhead ribozyme derived from schistosomes. *RNA* **11**, 187–196 (2005).
- Martick, M. *Structural Study of how the Hammerhead Ribozyme Solves the Problem of Catalysis*. PhD thesis, Univ. California, (<http://www.proquest.com/>) (2007).
- Yen, L. *et al.* Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* **431**, 471–476 (2004).
- Meaux, S. & Van Hoof, A. Yeast transcripts cleaved by an internal ribozyme provide new insight into the role of the cap and poly(A) tail in translation and mRNA decay. *RNA* **12**, 1323–1337 (2006).
- Lockyer, A. E. *et al.* The phylogeny of the Schistosomatidae based on three genes with emphasis on the interrelationships of Schistosoma Weinland, 1858. *Parasitology* **126**, 203–224 (2003).
- Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nature Rev. Mol. Cell Biol.* **8**, 113–126 (2007).
- Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
- Stern, S., Moazed, D. & Noller, H. F. Structural analysis of RNA using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol.* **164**, 481–489 (1988).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Hall and K. Chakrabarti for assistance with cell culture, members of Haussler and Ares laboratories for sharing tissue culture space, M. Robertson for the transcriptase and A. Zahler for discussion. This work was supported by the National Institutes of Health grant R01043393.

**Author Contributions** L.H.H. and M.M. did the sequence searches, designed the study, performed the experiments, analysed the data and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to M.M. ([mmartick@yahoo.com](mailto:mmartick@yahoo.com)).

## METHODS

**Motif search and secondary structure.** The RNABOB program<sup>11,12</sup> was used to search mRNA sequence data from the Ensembl database (release 48, <http://www.ensembl.org>) using a descriptor (Supplementary Fig. 1) for the hammerhead ribozyme motif. The secondary structure was built using information from the *Schistosoma* hammerhead ribozyme crystal structure<sup>3</sup>. Mouse and rat *Clec2* paralogues encoding partial and full ribozyme sequences were identified using BLAT and the UCSC Genome browser<sup>15,27</sup>. Alignment was generated using ClustalW<sup>29</sup>. *Clec2d* orthologues were identified using the 30-way Multiz alignment and conservation tool in the UCSC genome browser<sup>15</sup>. Sequences with similarity to stem 2 regions of the rodent CLEC2 ribozymes were selected for further analysis. These sequences were first screened for completeness of stem 2 and conserved hammerhead ribozyme nucleotides in the enzyme strand. The candidate ribozymes were then selected on the basis of the expected sequence (the ability to form stems 1 and 3 together with the enzyme strand) and orientation of the substrate within the 2,000 bp regions flanking stem 2.

**Constructs.** The hammerhead ribozyme-containing 3' UTR sequences (GenBank AF321553 nucleotides 684–1,179 for D-hcu; NM\_153506 nucleotides 695–1,710 for E-hcu) were cloned from total DNA (BioChain Institute, Inc.) of BALB/c mice. The DM1-hcu construct is the same as D-hcu with the following exception: the hammerhead ribozyme has been inactivated by a triple mutation spanning the cleavage site (5'-A(16.1)G(17)A(1.1)-3'). The DM2-hcu construct contains the hammerhead ribozyme inactivated by a single base (G8 to C8) mutation (refer to Fig. 1 for residue identification), with the rest of the sequence identical to D-hcu. The R-D-hcu construct contains D-hcu sequence in the antisense direction. The positive controls CHR (control hammerhead ribozyme) and CHRM (G8 to C8 mutant of CHR) were derived from *Schistosoma* ribozyme<sup>3</sup> and were oriented analogously to the CLEC2 ribozymes—with stem 3 connecting the structure to the rest of the transcript. Each of the two controls was capped with an artificial tetraloop closing stem 1 (full CHR sequence with tetraloop underlined and the CHRM mutation position preceded by an asterisk: 5'-CCGGCGT CCTGGTATCCAATCCTTCGGGATGTACTACCAGCTGAT\*GAGTCCCAAA TAGGACGAAACGCCGG-3'). For *in vitro* transcription assays, 3' UTR sequences were cloned into the XbaI site of the pET-20(b) vector (Novagen). For expression assays, 3' UTR sequences and control ribozymes were placed 7 nucleotides downstream of the firefly luciferase open reading frame (XbaI site) in psiCHECK-2 dual luciferase vector (Promega). Cognate poly A sites have been omitted from all 3' UTR sequences and replaced with vector-encoded SV40 Poly(A) signal. The psiCHECK-2 vector also contained a *Renilla* luciferase gene,

the expression product of which was used as an internal control for transfection efficiency and the extent of expression.

**In vitro transcription and reverse transcriptase primer extension.** pET-20(b) constructs were linearized and transcribed for 1 h at 37 °C using T7 RNAP in the following conditions: 50–60 µM double-stranded plasmid, 5 mM ribonucleotide triphosphates, 28 mM MgCl<sub>2</sub>, 2 mM spermidine-HCl, 100 mM dithiothreitol, 0.01% Triton X-100 and 40 mM Tris-HCl buffer (pH 8.0). Transcripts were purified using phenol and chloroform after the transcription reaction and then again after the treatment with RQ-DNase (Promega). Reverse transcriptase primer extensions and RNA sequencing were adapted for a [<sup>32</sup>P]-end-labelled primer from previously published methods<sup>28</sup>.

**Reporter expression assays.** NIH 3T3 mouse fibroblast cells were transiently transfected using Lipofectamine LTX (Invitrogen) according to the manufacturer's instructions and assayed using the Dual-Glo Luciferase Assay System (Promega). The firefly/*Renilla* luciferase expression ratios were normalized to those of the psiCHECK-2 vector.

**RT-PCR.** NIH 3T3 cells were transfected with mouse *Clec2d* 3' UTR-containing psiCHECK-2 plasmids in 10-cm plates. Whole-cell RNA was purified using Trizol (Sigma), treated with RQ-DNase (Promega), and purified with phenol and chloroform. cDNA was synthesized using two primers specific to two different (experimental and control) plasmid-derived transcripts using previous methods<sup>28</sup>. Primer sequences for cDNA synthesis were 5'-CATGTCTGCTCGAAGCGGC-3' for the ribozyme-containing firefly luciferase (*hLuc*<sup>+</sup>) transcript, and 5'-TCCTCAGGCTCCAGTTTCC-3' for the control *Renilla* luciferase (*hRLuc*) transcript. Primer extension products were treated with RNase A and RNase H, purified with phenol and chloroform, and resuspended in water. For PCR reactions, Accuprime Taq along with Buffer II (Invitrogen) were used according to manufacturer's instructions. For PCR of the ribozyme-containing *hLuc*<sup>+</sup>, the *hLuc*<sup>+</sup> cDNA primer was used together with 5'-GTGGACGAGGTGCCCCAAGG-3' (forward) for the 643 nucleotide product, and 5'-GCTACTCTTCCCCTATGGTC-3' (forward) for the 494 nucleotide product. For PCR of the control *hRLuc*, the *hRLuc* cDNA primer was used together with 5'-CTGATCTGATCGGAATGGG-3' (forward). PCR reactions were placed on ice after the designated number of cycles, resolved on a 2% agarose gel and stained with ethidium bromide. PCR product intensities were quantified using ImageQuant 5.2 (Molecular Dynamics).

29. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948 (2007).

# Dynamic thiolation–thioesterase structure of a non-ribosomal peptide synthetase

Dominique P. Frueh<sup>1</sup>, Haribabu Arthanari<sup>1</sup>, Alexander Koglin<sup>1</sup>, David A. Vosburg<sup>1†</sup>, Andrew E. Bennett<sup>1</sup>, Christopher T. Walsh<sup>1</sup> & Gerhard Wagner<sup>1</sup>

Non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) produce numerous secondary metabolites with various therapeutic/antibiotic properties<sup>1</sup>. Like fatty acid synthases (FAS), these enzymes are organized in modular assembly lines in which each module, made of conserved domains, incorporates a given monomer unit into the growing chain. Knowledge about domain or module interactions may enable reengineering of this assembly line enzymatic organization and open avenues for the design of new bioactive compounds with improved therapeutic properties. So far, little structural information has been available on how the domains interact and communicate. This may be because of inherent interdomain mobility hindering crystallization, or because crystallized molecules may not represent the active domain orientations<sup>2</sup>. In solution, the large size and internal dynamics of multidomain fragments (>35 kilodaltons) make structure determination by nuclear magnetic resonance a challenge and require advanced technologies. Here we present the solution structure of the apo-thiolation–thioesterase (T–TE) di-domain fragment of the *Escherichia coli* enterobactin synthetase EntF NRPS subunit. In the holoenzyme, the T domain carries the growing chain tethered to a 4'-phosphopantetheine whereas the TE domain catalyses hydrolysis and cyclization of the iron chelator enterobactin. The T–TE di-domain forms a compact but dynamic structure with a well-defined domain interface; the two active sites are at a suitable distance for substrate transfer from T to TE. We observe extensive interdomain and intradomain motions for well-defined regions and show that these are modulated by interactions with proteins that participate in the biosynthesis. The T–TE interaction described here provides a model for NRPS, PKS and FAS function in general as T–TE-like di-domains typically catalyse the last step in numerous assembly-line chain-termination machineries.

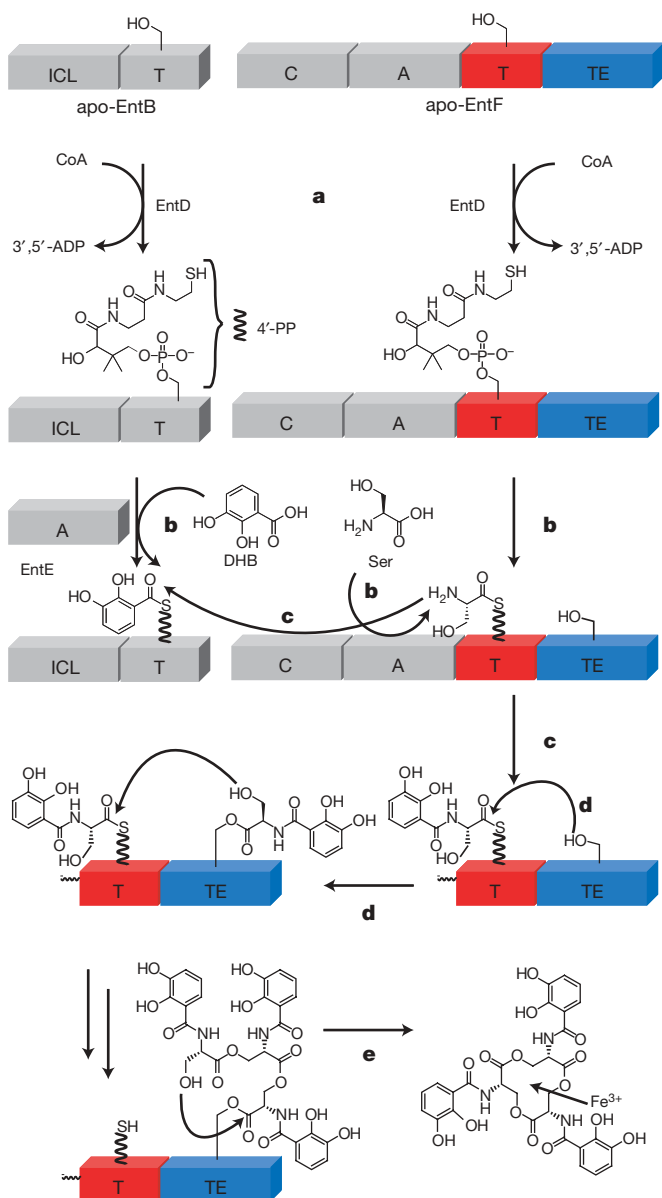
The *Escherichia coli* enterobactin synthetase (Ent) synthesizes the siderophore enterobactin, a virulence factor used by *E. coli* to infect iron-limited microenvironments of vertebrate hosts. Enterobactin is obtained by three iterative condensations of dihydroxybenzoate (DHB) with serine, followed by macrocyclizing release of the macro-lactone that can form high-affinity hexadentate complexes with ferric iron<sup>3,4</sup> (Fig. 1). Ent is a two-module (EntB/EntE and EntF), three-protein assembly line that combines the features of both a type I NRPS, in which all domains interact *in cis* in a single protein chain (EntF), and a type II NRPS, in which the domains interact *in trans* as free-standing units (EntB/EntE). The dedicated phosphopantetheinyl transferase (PPTase), EntD, loads the apo form of the T domain with 4'-phosphopantetheine (4'-PP), which tethers the growing acyl chain while allowing the bound substrates to navigate between the catalytic sites (Fig. 1a). The EntE and EntF adenylation (A) domains select the appropriate substrates (DHB or serine) and activate them

as acyl-adenosine-monophosphates. These are then loaded on the 4'-PP arms attached to Ser 245 of the holo-EntB aryl carrier protein (ArCP) and to Ser 48 of holo-EntF T domains, respectively (Fig. 1b). The EntF condensation (C) domain catalyses the DHB–Ser amide-bond formation (Fig. 1c). The DHB–Ser chains are next transferred to the active site Ser 180 of the thioesterase (TE) domain, thereby making the 4'-PP arm of the T domain available for the next cycle of condensation (Fig. 1d). In this step of the reaction, the 4'-PP arm (wavy line in Fig. 1d) has to reach from Ser 48 of the T domain to Ser 180 of the TE domain, in which the triad Ser 180, Asp 207 and His 313 catalyses the oxoester bond formation between the substrate and Ser 180. The steps shown in Fig. 1b–d are repeated twice to produce the tethered linear DHB–Ser trimer before its intramolecular cyclization by the TE domain of EntF, which constitutes the release of the enterobactin trilactone (Fig. 1e). The structural basis of communication between the T and TE domains in chain-release steps at the end of NRPS assembly lines has been so far unknown.

Here we describe the structure of the 37 kDa apo-EntF T–TE di-domain in which the active site Ser 48 (of the T domain) was replaced with Ala to enable production of a homogeneous sample. This mutation is isomorphous as the NMRs of the T–TE di-domain are unchanged except for the immediate environment of residue 48, and the interdomain contacts are identical (see Methods and Supplementary Figs 1 and 2). It represents the physiological structure before CoA addition as indicated at the top of Fig. 1a. The mutated apoprotein allows us to distinguish domain interactions from domain/substrate contributions (including the 4'-PP arm), a crucial knowledge for successful assembly line reprogramming by domain swap. The T and TE domains have a well-defined relative orientation with a primarily hydrophobic interface (Fig. 2a, c). This results in a 1,300 Å<sup>2</sup> buried surface area not including the linker, which interacts with both domains and contributes to a further 200 Å<sup>2</sup>. The globular T domain forms a three-helical bundle that is wedged between the globular core of the thioesterase and two  $\alpha$ -helices ( $\alpha_{4TE}$ – $\alpha_{5TE}$ ) protruding from this core. These helices resemble two webbed fingers that form a lid covering the T and TE active sites. Ser48Ala and Ser 180 are found 17 Å apart at both ends of a canyon formed by loop L12<sub>TE</sub> and helix  $\alpha_{4TE}$ . This allows the 4'-PP arm, which is tethered to Ser 48 of the T domain in the holoprotein and can span up to ~20 Å, to reach Ser 180 of the TE domain. Dynamics data indicate that the two helices ( $\alpha_{4TE}$ – $\alpha_{5TE}$ ) form a mobile flap that is able to open to accommodate the 4'-PP arm (see below). Thus, the structure should not be interpreted as rigid but rapidly fluctuating to allow loading of the 4'-PP arm and the growing enterobactin chain. This resembles many other proteins that have to undergo major opening fluctuations to allow binding of target proteins or substrates, such as HIV protease<sup>5</sup>.

<sup>1</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>†</sup>Present address: Department of Chemistry, Harvey Mudd College, 301 Platt Boulevard, Claremont, California 91711-5901, USA.

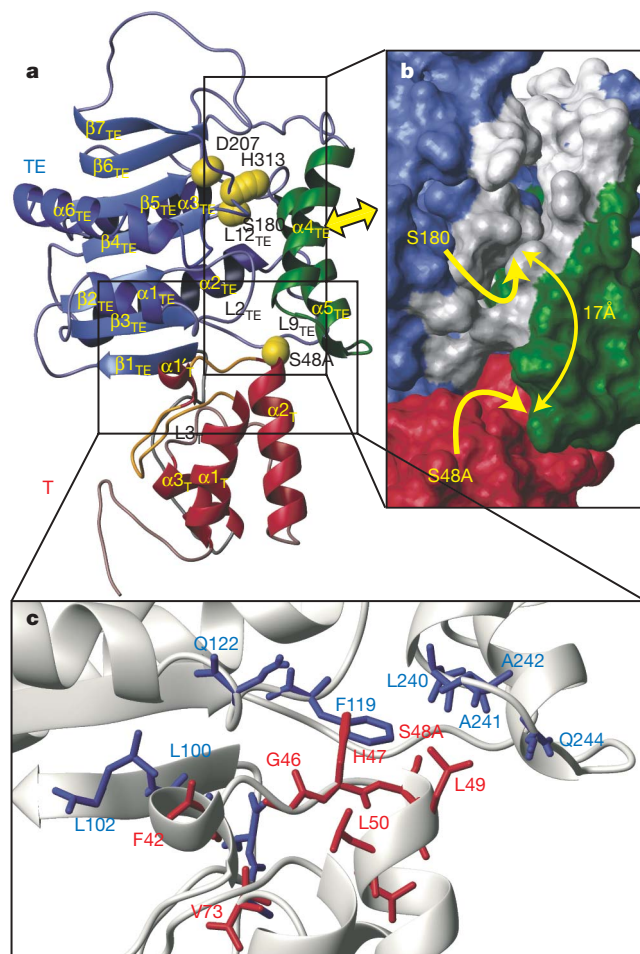
The T domain contacts TE via the beginning of helix  $\alpha_{2T}$  that contains Ser48Ala, the end of loop L3<sub>T</sub>, and the single turn helix  $\alpha_{1T}$  (Fig. 2c). The region near Ser48Ala interacts with both the core of the TE domain, via L2<sub>TE</sub>, and the tips of the helical fingers, which thus cover the T active site (Fig. 2b). In EntF, helix  $\alpha_{1T}$  stabilizes the hydrophobic core of the T domain with a phenylalanine side chain (Phe41) anchored in the heart of the helical bundle. An adjacent phenylalanine (Phe 42) interacts with hydrophobic residues of  $\beta_{1TE}$ . Point mutations of either phenylalanines lead to a disruption of the T/TE domain interaction and an unfolded T domain (data not shown).  $\alpha_{1T}$  is absent in the only other structure of an (excised) T domain<sup>6,7</sup>, but has been observed in related FAS and PKS acyl carrier proteins (ACP) and may contribute to domain recognition<sup>8–10</sup> (see Supplementary Information). Loop L3<sub>T</sub>, which interacts with  $\beta_{1TE}$ , features helical characteristics and is mobile as indicated by the low intensities of the corresponding NMR signals (Supplementary Fig. 8).



**Figure 1 | Summary of enterobactin synthesis.** The grey, red and blue boxes represent individual domains; the C, A, T and TE domains are described in the text. The EntB ICL domain is an isochorismate ligase that participates in the conversion of chorismate to DHB, a preliminary step to the synthesis of enterobactin. Initially, EntD loads the phosphopantetheinyl arm (4'-PP, represented as a wavy line) onto the active-site serines of apo-EntB and apo-EntF. The synthesis steps (a–e) are described in the text.

In general, the T domain shows internal mobility as further supported by fast NH exchange (Supplementary Fig. 9). Such dynamics have been observed for single-domain ACP-type T domains in FAS systems<sup>11</sup>, and may be a general property of acyl and peptidyl carrier proteins (see Supplementary Information).

The TE domain adopts the overall fold of  $\alpha/\beta$ -hydrolases. However, the 'finger' region differs in the two systems for which TE structures have been determined<sup>12,13</sup>. Helix  $\alpha_{4TE}$  faces the catalytic triad, located in a conserved region, and delimits part of the 'cyclization bucket' previously described for substrate macrocyclization in surfactin type I thioesterase<sup>12</sup> (TEI; Fig. 2b, grey and white). For surfactin TEI, two conformers have been observed for this helix<sup>12</sup>, whereas the electron density was too low in fengimycin TEI to allow any structural interpretation<sup>13</sup> (see Supplementary Information). Moreover, NMR studies of the free standing surfactin type II thioesterase (TEII) indicated two conformations in the corresponding lid<sup>14</sup>. Accordingly, many residues in this region of the EntF TE domain show reduced intensities in their NMR signals, which is indicative of slow modulations of their environments (Supplementary Fig. 8). Furthermore, unlike the core of the TE domain, all amide protons of the finger region exchange rapidly with



**Figure 2 | Structure of the EntF T-TE fragment.** a, A ribbon diagram is shown. The T domain (red) is wedged between the core (blue) and the lid (green) of the TE domain. Active sites are shown as yellow spheres. The double-headed arrow emphasizes that the flap formed by the  $\alpha_{4TE}$  and  $\alpha_{5TE}$  helices is relatively mobile, opening frequently, which seems to be necessary to accommodate the 4'-PP arm in the processes depicted in Fig. 1. b, Surface representation of the region containing the canyon (grey) which must open to accommodate the 4'-PP arm in the holoprotein. The cyclization bucket<sup>12</sup> is shown in white and grey. In this conformation, Ser48Ala and Ser 180 are buried. c, The domain interface is shown. The side chains of all residues giving rise to interdomain nuclear Overhauser effects (NOEs) are shown.

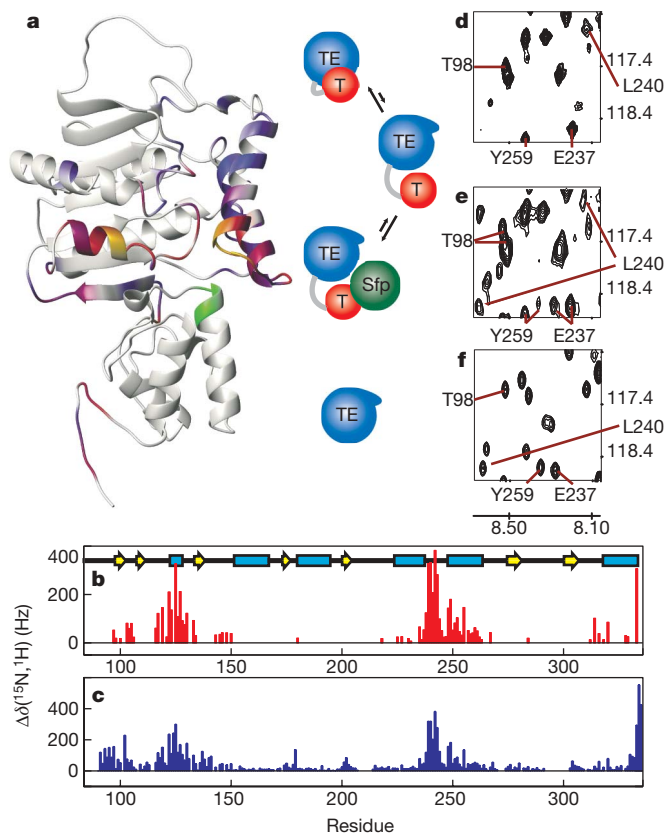
solvent (Supplementary Fig. 9), indicative of breathing/opening motions. Such mobility may be necessary to allow for access to both T and TE active sites, which would be occluded in a more rigid canyon structure (see later). The mobility also provides conformational plasticity to the TE active site to accommodate the various intermediates stored on Ser 180 and allow the 4'-PP arm to navigate in and out of the canyon in the holoprotein, which would not be possible in a static interpretation of our solution structure.

In addition to the intradomain dynamics reported above, we observed an open form of the T-TE molecule, T-TE<sup>0</sup>, in which TE and T interact little or not at all (see Supplementary Information and Fig. 3). Indeed, although ACP-type T domains have recently been observed in FAS<sup>15,16</sup>, they have previously been reported to be invisible in the crystal structures of FAS or PKS due to low electron density<sup>17-19</sup>, in agreement with dynamic domain interactions. To investigate the significance of these dynamic events, the T-TE fragment was titrated with binding partners that participate in different steps of enterobactin synthesis. To become functional as carrier proteins, T domains first need to be primed with 4'-PP arms by a PPTase (Fig. 1a). When the broad-specificity PPTase Sfp<sup>20</sup> was added to a solution of T-TE, a binding site was identified around the active-site residue 48, as expected<sup>6</sup> (Fig. 3a, green). In addition, however, the subset of signals of the T-TE<sup>0</sup> conformer increased, suggesting that Sfp drives the equilibrium towards the dislodged form of the di-domain (Fig. 3). This is reasonable because Ser48Ala is partially covered by the tips of the fingers in the di-domain (Fig. 2b). The

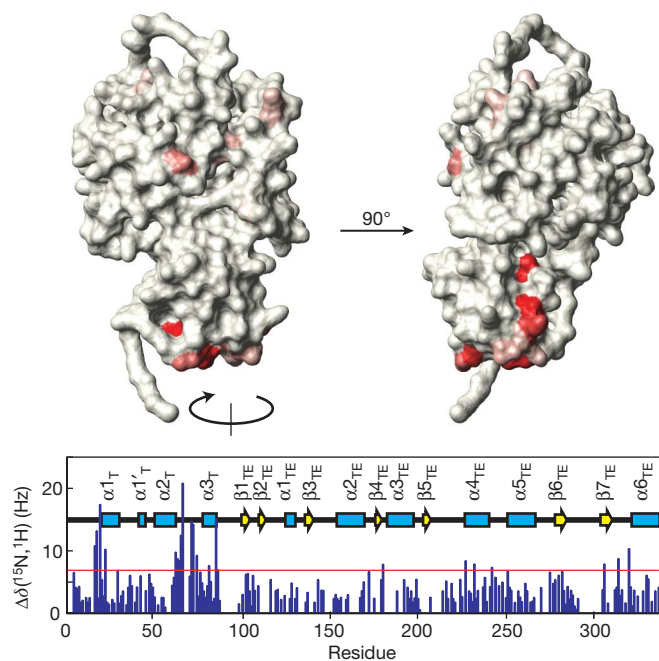
simultaneous observation of new signals and shifts of resonances (Supplementary Fig. 11) indicates the presence of complex dynamic processes involved in the function of this enzyme. The EntF-specific EntD enzyme<sup>20</sup> caused similar selection of the open state; in contrast, the ACP-specific AcpS PPTase<sup>21</sup> did not (Supplementary Fig. 12), highlighting the specificity of the mechanism.

The EntF C domain<sup>22</sup> condenses the substrates that are covalently attached as thioesters to 4'-PP arms on the EntB and EntF T domains (Fig. 1c). Specific chemical shift changes on addition of the EntF C domain to T-TE showed the C-domain-binding face on T (Fig. 4), and only minor effects were observed on TE. Because the T/C surface does not overlap with the TE interaction surface, no disruption of the T/TE interface is necessary and only a modulation of their interaction is observed. Notably, no interaction was observed when the TycC5 (the fifth module of the tyrocidine synthetase C subunit) C domain was added to a T-TE solution, suggesting that this T-C recognition is EntF specific (see Supplementary Fig. 13). For both C-domain and Sfp interactions, perturbations on TE are not limited to the interface, but affect the finger region and the TE active site, showing that the binding events probably influence the dynamics and/or the conformations in these areas. These observations apply to the apo form of the molecule and thus only report on domain-domain interactions. It is probable that during biosynthesis the 4'-PP arm, loaded with various intermediates, also affects these dynamic effects.

The structure of the T-TE di-domain provides a starting point for elucidating the communication between domains in NRPS systems. Understanding how the central carrier domains interact with their partner catalytic domains is a prerequisite for understanding the NRPS assembly logic in chain initiation, elongation and termination steps. Our data document a well-defined, yet dynamic, interface between the T and TE domains, consistent with the idea that the T domain must interact with the upstream C and A domains as well as



**Figure 3 | Interaction with Sfp.** **a**, Residues subject to two different environments. The colour code represents small (blue), medium (red) and large (yellow) chemical shift differences between the two forms as reported in **b**, **c**. Chemical shift differences between the major conformer of free T-TE and the excised TE domain. **d-f**, Spectral signatures of the Sfp interaction. **d**, The reference spectrum of Ser48Ala. The open form is only visible with longer acquisitions (Supplementary Fig. 3). On addition of Sfp, the signals of the residues in green (see **a**) disappear from the spectrum, and those of the second form appear (**e**). **f**, Corresponding region of the spectrum of the excised TE.



**Figure 4 | Interaction with the C domain.** Many resonances experience shifts on addition of the C domain (bottom), indicating a modulation of the environment of the corresponding residues. Those belonging to the T domain form an interaction surface that does not overlap with the T-TE interface (top). Thus, no disruption of the T-TE interaction is observed. The effects on the TE domain may be due to weak, indirect interactions with the C domain, or to a secondary effect due to fluctuations of the dynamics or modifications of the conformations in these regions. Residues with shifts larger than one s.d. from the mean (red line, bottom) are colour coded with a gradation from white to red.

this downstream TE domain. The PPTase and C domain, both of which must visit the T domain, modulate the interaction between the T and TE domains as well as the conformational and/or dynamic properties of each domain. Notably, the *in trans* interaction between surfactin TEII and the TycC3 T domain uses a different domain orientation but places the two active sites at a comparable distance<sup>14</sup>. Moreover, the *in trans* domain interaction selects subsets of conformational states similarly to the *in cis* situation described here. NRPS domains need to communicate with each other at specific points during the peptide synthesis, hence natural product assembly lines must be dynamic entities. The observed malleability may serve to regulate this interplay by stabilizing a given interaction during the course of the corresponding reaction.

## METHODS SUMMARY

Recombinant protein was produced in an *E. coli* expression system. The Ser48Ala mutation was used because it was difficult to produce exclusively holo- or apo-T-TE, rather than a mixture, in bacterial expression systems tested. Moreover, wild-type apo-T-TE exists in two inter-converting conformational states for the T domain (data not shown); a similar phenomenon is observed for the apo-T domain of tyrocidin synthetase, TycC3, in which the Ser48Ala mutation locks the protein in a single conformation<sup>6</sup>. Out of the 337 backbone resonances and the associated side-chain resonances, 313 were assigned with transverse relaxation optimized spectroscopy (TROSY) versions of conventional experiments<sup>23</sup> sometimes with non-uniform sampling<sup>24</sup>. A double TROSY hNcaNH was recorded to facilitate the assignment<sup>25</sup>. Most distance restraints were obtained from time-shared nuclear Overhauser enhancement spectroscopy (NOESY) experiments<sup>26</sup>. High-resolution structures were obtained with the program CYANA<sup>27</sup> (backbone root mean square deviations (r.m.s.d.) of 1.2 Å for residues in secondary structure elements, with corresponding r.m.s.d. for the single domains at 0.6 Å (T domain) and 1.0 Å (TE domain; Fig. 2 and Supplementary Fig. 3)). The presence of complex dynamic process prevented the use of residual dipolar couplings<sup>28,29</sup>. Titrations were conducted with a constant T-TE concentration and with decreasing concentration of the titrant, to ensure that no degradation was induced by the titration. The determination of solvent accessible residues was obtained by changing the solvent from H<sub>2</sub>O to D<sub>2</sub>O with a PD10 column (GE healthcare).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 10 September 2007; accepted 10 June 2008.

- Walsh, C. T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**, 1805–1810 (2004).
- Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L. O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781–792 (2007).
- Gehring, A. M., Bradley, K. A. & Walsh, C. T. Enterobactin biosynthesis in *Escherichia coli*: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* **36**, 8495–8503 (1997).
- Gehring, A. M., Mori, I. & Walsh, C. T. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37**, 2648–2659 (1998).
- Wagner, G. The importance of being floppy. *Nature Struct. Biol.* **2**, 255–257 (1995).
- Koglin, A. *et al.* Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* **312**, 273–276 (2006).
- Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A. & Holak, T. A. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* **8**, 407–418 (2000).
- Findlow, S. C., Winsor, C., Simpson, T. J., Crosby, J. & Crump, M. P. Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from *Streptomyces rimosus*. *Biochemistry* **42**, 8423–8433 (2003).
- Johnson, M. A., Peti, W., Herrmann, T., Wilson, I. A. & Wuthrich, K. Solution structure of Asl1650, an acyl carrier protein from *Anabaena* sp. PCC 7120 with a variant phosphopantetheinylation-site sequence. *Protein Sci.* **15**, 1030–1041 (2006).

- Zornetzer, G. A., Fox, B. G. & Markley, J. L. Solution structures of spinach acyl carrier protein with decanoate and stearate. *Biochemistry* **45**, 5217–5227 (2006).
- Kim, Y. & Prestegard, J. H. A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry* **28**, 8792–8797 (1989).
- Bruner, S. D. *et al.* Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**, 301–310 (2002).
- Samel, S. A., Wagner, B., Marahiel, M. A. & Essen, L. O. The thioesterase domain of the fengycin biosynthesis cluster: a structural base for the macrocyclization of a non-ribosomal lipopeptide. *J. Mol. Biol.* **359**, 876–889 (2006).
- Koglin, A. *et al.* Structural basis for the selectivity of the external thioesterase of the surfactin synthetase. *Nature* doi:10.1038/nature07161 (this issue).
- Leibundgut, M., Jenni, S., Frick, C. & Ban, N. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* **316**, 288–290 (2007).
- Lomakin, I. B., Xiong, Y. & Steitz, T. A. The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell* **129**, 319–332 (2007).
- Maier, T., Jenni, S. & Ban, N. Architecture of mammalian fatty acid synthase at 4.5 Å resolution. *Science* **311**, 1258–1262 (2006).
- Jenni, S., Leibundgut, M., Maier, T. & Ban, N. Architecture of a fungal fatty acid synthase at 5 Å resolution. *Science* **311**, 1263–1267 (2006).
- Jenni, S. *et al.* Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254–261 (2007).
- Lambalot, R. H. *et al.* A new enzyme superfamily—the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–936 (1996).
- Lambalot, R. H. & Walsh, C. T. Cloning, overproduction, and characterization of the *Escherichia coli* holo-acyl carrier protein synthase. *J. Biol. Chem.* **270**, 24658–24661 (1995).
- Roche, E. D. & Walsh, C. T. Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry* **42**, 1334–1344 (2003).
- Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wuthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl Acad. Sci. USA* **95**, 13585–13590 (1998).
- Rovnyak, D. *et al.* Accelerated acquisition of high resolution triple-resonance spectra using non-uniform sampling and maximum entropy reconstruction. *J. Magn. Reson.* **170**, 15–21 (2004).
- Frueh, D. P. *et al.* Non-uniformly sampled double-TROSY hNcaNH experiments for NMR sequential assignments of large proteins. *J. Am. Chem. Soc.* **128**, 5757–5763 (2006).
- Frueh, D. P., Vosburg, D. A., Walsh, C. T. & Wagner, G. Determination of all NOEs in <sup>1</sup>H–<sup>13</sup>C-Me-ILV-U-<sup>2</sup>H–<sup>15</sup>N proteins with two time-shared experiments. *J. Biomol. NMR* **34**, 31–40 (2006).
- Guntert, P., Mumenthaler, C. & Wuthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).
- Tjandra, N. & Bax, A. Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114 (1997).
- Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc. Natl Acad. Sci. USA* **92**, 9279–9283 (1995).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Sastry and A. Marintchev for advice in sample preparation; Z. Zhou for providing the Ser48Ala plasmid; Z.-Y. Sun for participating in the development of new NMR techniques; and P. Selenko for comments on the manuscript. This work was supported by NIH grants GM47467, EB 002026, and GM066360 and a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund (D.A.V.). A.K. thanks the Human Frontier Science Program for a long-term fellowship awarded in April 2007.

**Author Contributions** D.P.F., G.W. and C.T.W. designed research and wrote the manuscript. D.P.F. conducted the research including protein expression, resonance assignment and structure calculation. H.A. prepared various samples and helped during the acquisition of NMR experiments. D.A.V. prepared various samples. A.K. helped during the structure calculation and prepared the C domains EntD and Sfp samples. A.E.B. participated in early stages of the project.

**Author Information** Atomic coordinates have been deposited in the Protein Data Bank under accession code 2ROQ. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to G.W. ([gerhard\\_wagner@hms.harvard.edu](mailto:gerhard_wagner@hms.harvard.edu)), C.T.W. ([christopher\\_walsh@hms.harvard.edu](mailto:christopher_walsh@hms.harvard.edu)), or D.P.F. ([dominique\\_frueh@hms.harvard.edu](mailto:dominique_frueh@hms.harvard.edu)).

## METHODS

**Protein expression and purification.** A pET30a+ plasmid containing the EntF T–TE gene, with a C-terminal His<sub>6</sub> tag, was transformed into *E. coli* BL21(DE3) cells. Isotope enrichment of protonated samples was produced by overexpression in M9 minimal media in H<sub>2</sub>O containing <sup>15</sup>NH<sub>4</sub>Cl or <sup>13</sup>C glucose or both. Deuterated samples (including methyl labelled samples) were prepared in D<sub>2</sub>O containing <sup>2</sup>H-<sup>13</sup>C glucose. Partially deuterated samples were prepared in recycled D<sub>2</sub>O with <sup>13</sup>C glucose. ‘ILV methyl labelled’ samples, which are selectively protonated on methyls of Leu, Val and Ile (δ1), were prepared as described in<sup>26,30</sup>, with α-ketobutyrate and α-ketoisovalerate precursors either uniformly enriched in <sup>13</sup>C, for assignments of side-chain signals, or only <sup>13</sup>C enriched at the methyls of the precursors, for the NOE experiments. Samples selectively enriched in <sup>15</sup>N for a given amino acid were prepared by supplying the medium with a single <sup>15</sup>N-enriched amino acid and 19 unlabelled amino acids<sup>31</sup>. The growth times were 4 h in H<sub>2</sub>O, 6 h in recycled D<sub>2</sub>O and 12 h in D<sub>2</sub>O. Induction times were 2 h in H<sub>2</sub>O, 4 h in recycled D<sub>2</sub>O and 6 h in D<sub>2</sub>O. For all samples, the cells were lysed by sonication and the protein was purified using Ni–NTA resin (Qiagen) followed by FPLC using a Sephadex gel-filtration column (S75 or S200). The final sample concentrations were approximately 300 μM in 20 mM phosphate buffer (pH = 6.7) with 150 mM NaCl, and 1 mM EDTA and dithiothreitol.

**Assignment of backbone resonances.** TROSY versions of backbone experiments<sup>23,32,33</sup> were used in a 70% deuterated <sup>2</sup>H-<sup>15</sup>N-<sup>13</sup>C sample. This led to the assignment of 50% of the signals, 30% of which were assigned with the automated assignment program IBIS<sup>34</sup>. HN-TROSY-HSQC<sup>35</sup> spectra were recorded on samples selectively enriched in <sup>15</sup>N for Ile, Leu, Val, Lys, Phe, Tyr and Gly (that partially scrambles into Ser). These spectra were combined with the <sup>1</sup>H, <sup>15</sup>N projections of a CCONH spectrum of a ILV methyl-labelled sample and with the <sup>1</sup>H, <sup>15</sup>N projections of an HNCOCACB spectrum. These combinations enabled the identification of Ile, Leu, Val, Lys, Phe, Tyr, Gly and Ser residues that were neighbours of Ile, Leu, Val and Gly residues. Similarly, a non-uniformly sampled<sup>24</sup> HNCO spectrum was recorded on a sample uniformly labelled in <sup>15</sup>N and selectively labelled on carbonyl carbons of prolines. In addition, a non-uniformly sampled double-TROSY-hNcaNH experiment<sup>25</sup> was designed to correlate triplets of amide nitrogens belonging to contiguous residues. All of the information together led to 70% of the backbone assignment. A different strategy was used to assign residues with non-exchangeable amide protons, which constitute 30% of the TE domain. Non-uniform sampling was used to record high-resolution HNCO, HNCA and HNCOCACB spectra of a <sup>15</sup>N-<sup>13</sup>C-labelled sample at 600 MHz with a good signal-to-noise ratio in spite of the fast relaxation due to the sample protonation. A <sup>13</sup>C-detected MQ-HACACO<sup>36</sup> was then used to assign Hα signals. These experiments were used together with a <sup>15</sup>N-dispersed NOESY-TROSY-HSQC leading to the assignment of 93% of the backbone resonances (313 out of 337 residues).

**Assignment of side-chain resonances.** HCCONH and CCONH<sup>37</sup> experiments were recorded on a 70% deuterated, uniformly <sup>15</sup>N- and <sup>13</sup>C-enriched sample. These experiments were also recorded on a ILV-methyl-labelled sample together with HCCH-total correlation spectroscopy (TOCSY)<sup>38</sup>. An HCCH-COSY<sup>39</sup> was recorded on a <sup>15</sup>N-<sup>13</sup>C-enriched sample to assign the signals of residues with non-exchangeable amide protons. Spin systems obtained with this experiment were related to backbone assignments by the MQ-HACACO experiment. Further assignments were obtained with NOESY spectra described below, after initial structure calculations.

**Measurement of NMR constraints.** Torsion angles constraints were obtained with the program TALOS<sup>40</sup>. Most of the distance constraints were measured with time-shared three-dimensional (3D) TS-NOESY-HN-TROSY/NOESY-HC-PEP-HSQC and 4D HN-TROSY/HC-HSQC-NOESY-HN-TROSY/HC-PEP-HSQC experiments<sup>26</sup> with a mixing time (τ<sub>m</sub>) of 200 ms, recorded on a ILV-methyl-labelled sample that also contained protonated phenylalanine<sup>41</sup>. A time-shared HSQC-NOESY (D.P.F., unpublished data), with high resolution for the dimension featuring the cross-peaks, was recorded to provide accurate integration of the NOE signals. A NOESY-HN-TROSY spectrum (τ<sub>m</sub> = 90 ms) was recorded on a protonated sample to measure restraints involving non-exchangeable amide protons.

**NMR data acquisition.** All experiments were conducted on Bruker spectrometers equipped with cryoprobes. For the assignment of backbone resonances, experiments were recorded on a 900 MHz spectrometer, except for those involving <sup>13</sup>C-<sup>15</sup>N samples and selectively labelled samples, which were measured at 600 MHz, and the MQ-HACACO experiment, which was recorded on a 500 MHz spectrometer equipped with a TXI cryoprobe. Side-chain experiments were measured at 600 MHz, except for the HCCH-COSY, which was recorded at 900 MHz. NOESY spectra were recorded at 750 MHz (3D and 4D time-shared experiments) and 900 MHz (3D time-shared HSQC-NOESY and 3D NOESY-TROSY). All spectra were processed with nmrPipe<sup>42</sup> and analysed with CARRA<sup>43</sup>.

**Structure calculation.** The structures of each domain were first solved individually, with constraints measured on the T–TE fragment. For the T domain, a homology model with the TycC3 T domain (PDB code 1JMK<sup>20</sup>) could be produced<sup>44</sup>. For the TE domain, sequential alignment with the surfactin TEI (1DNY<sup>12</sup>) provided erroneous results. However, we could use secondary structure alignment, on the basis of TALOS predictions and NOE patterns, to produce a model with the program DeepView<sup>44,45</sup>. The program CYANA<sup>27</sup> was used to predict NOEs on the basis of these models. The quality of the model could immediately be inspected in the various NOESY spectra and numerous corrections were made, in particular with help from 4D spectra. This resulted in a first low-resolution structure (r.m.s.d. of about 4 Å in each domain and 10 Å for the T–TE fragment). This structure was then used to identify new distance constraints and to make further corrections. This was repeated for each new structure. One-hundred-and-eighty-one hydrogen bond constraints were added for residues in secondary structured regions, once the local r.m.s.d. were shown to be under 0.5 Å, and 497 torsion angle constraints were included when they supported the existing structure. In the end, 4,383 distance constraints (1,163 intraresidues, 2,546 medium range and 935 long range) with 34 interdomain distance constraints were used to produce a structure with a global r.m.s.d. of 2.2 Å. The structural refinement in explicit water was performed using the RECOORD<sup>46</sup> scripts, after converting the CYANA derived structures into the CNS/XPLOR<sup>47</sup> format.

**Titration.** See Supplementary Information. The reported chemical shift differences were calculated by  $\Delta\delta(^1\text{H}, ^{15}\text{N}) = ((\Delta\omega_{\text{H}})^2 + ((1/10)\Delta\omega_{\text{N}})^2)^{1/2}$ , where  $\Delta\omega_i$  is the chemical shift difference between the two species for nucleus *i*. A red dashed line on the resulting figures indicates a threshold of 1 s.d. that was used to single out outliers.

30. Gardner, K. H. & Kay, L. E. Production and incorporation of <sup>15</sup>N, <sup>13</sup>C, <sup>2</sup>H (<sup>1</sup>H-δ1 methyl) isoleucine into proteins for multidimensional NMR studies. *J. Am. Chem. Soc.* **119**, 7599–7600 (1997).
31. Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E. & Dahlquist, F. W. Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance. *Methods Enzymol.* **177**, 44–73 (1989).
32. Ferentz, A. E. & Wagner, G. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **33**, 29–65 (2000).
33. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).
34. Hyberts, S. G. & Wagner, G. IBIS – a tool for automated sequential assignment of protein spectra from triple resonance experiments. *J. Biomol. NMR* **26**, 335–344 (2003).
35. Pervushin, K., Riek, R., Wider, G. & Wuthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl Acad. Sci. USA* **94**, 12366–12371 (1997).
36. Pervushin, K. & Eletsky, A. A new strategy for backbone resonance assignment in large proteins using a MQ-HACACO experiment. *J. Biomol. NMR* **25**, 147–152 (2003).
37. Grzesiek, S., Anglister, J. & Bax, A. Correlation of backbone amide and aliphatic side-chain resonances in <sup>13</sup>C/<sup>15</sup>N-enriched proteins by isotropic mixing of <sup>13</sup>C magnetization. *J. Magn. Reson. B* **101**, 114–119 (1993).
38. Bax, A., Clore, M. & Gronenborn, A. M. <sup>1</sup>H-<sup>1</sup>H Correlation via isotropic mixing of <sup>13</sup>C magnetization, a new three-dimensional approach for assigning <sup>1</sup>H and <sup>13</sup>C spectra of <sup>13</sup>C-enriched proteins. *J. Magn. Reson.* **88**, 425–431 (1990).
39. Ikura, M., Kay, L. E. & Bax, A. Improved three-dimensional <sup>1</sup>H-<sup>13</sup>C-<sup>1</sup>H correlation spectroscopy of a <sup>13</sup>C-labeled protein using constant-time evolution. *J. Biomol. NMR* **1**, 299–304 (1991).
40. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
41. Medek, A., Olejniczak, E. T., Meadows, R. P. & Fesik, S. W. An approach for high-throughput structure determination of proteins by NMR spectroscopy. *J. Biomol. NMR* **18**, 229–238 (2000).
42. Delaglio, F. et al. NMRPipe: A Multidimensional Spectra Processing System Based on UNIX Pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
43. Keller, R. L. *The Computer Aided Resonance Assignment Tutorial* (Cantina, 2004).
44. Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381–3385 (2003).
45. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997).
46. Nederveen, A. J. et al. RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**, 662–672 (2005).
47. Brünger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

# Structural basis for the selectivity of the external thioesterase of the surfactin synthetase

Alexander Koglin<sup>1,2</sup>, Frank Löhr<sup>1</sup>, Frank Bernhard<sup>1</sup>, Vladimir V. Rogov<sup>1,3</sup>, Dominique P. Frueh<sup>2</sup>, Eric R. Strieter<sup>2</sup>, Mohammad R. Mofid<sup>4</sup>, Peter Güntert<sup>1,5</sup>, Gerhard Wagner<sup>2</sup>, Christopher T. Walsh<sup>2</sup>, Mohamed A. Marahiel<sup>4</sup> & Volker Dötsch<sup>1</sup>

**Non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS) found in bacteria, fungi and plants use two different types of thioesterases for the production of highly active biological compounds<sup>1,2</sup>. Type I thioesterases (TEI) catalyse the release step from the assembly line<sup>3</sup> of the final product where it is transported from one reaction centre to the next as a thioester linked to a 4'-phosphopantetheine (4'-PP) cofactor that is covalently attached to thiolation (T) domains<sup>4-9</sup>. The second enzyme involved in the synthesis of these secondary metabolites, the type II thioesterase (TEII), is a crucial repair enzyme for the regeneration of functional 4'-PP cofactors of holo-T domains of NRPS and PKS systems<sup>10-12</sup>. Mispriming of 4'-PP cofactors by acetyl- and short-chain acyl-residues interrupts the biosynthetic system. This repair reaction is very important, because roughly 80% of CoA, the precursor of the 4'-PP cofactor, is acetylated in bacteria<sup>13</sup>. Here we report the three-dimensional structure of a type II thioesterase from *Bacillus subtilis* free and in complex with a T domain. Comparison with structures of TEI enzymes<sup>3,14</sup> shows the basis for substrate selectivity and the different modes of interaction of TEII and TEI enzymes with T domains. Furthermore, we show that the TEII enzyme exists in several conformations of which only one is selected on interaction with its native substrate, a modified holo-T domain.**

The cyclic lipopeptide surfactin is one of the most potent bio-surfactants showing antibacterial and antiviral activity<sup>2,3,15</sup>. It is synthesized by a complex of three large surfactin synthetase subunits that consist of either three modules (subunits SrfA-A and SrfA-B) or one module (subunit SrfA-C; Fig. 1), with each module being responsible for the addition of one amino acid<sup>2,3</sup>. The transport of the growing chain between individual modules is achieved by small ~80-amino-acid-long T domains that interact with the adenylation (A) domain and both the upstream and downstream peptide-bond-forming condensation (C) domains<sup>4-9,16</sup>. In addition, some of the T domains of the surfactin synthetase interact with epimerization domains, located at the carboxy termini of the subunits SrfA-A and SrfA-B, or with the covalently linked TEI<sup>3</sup>. The TEI catalyses the macrolactone formation between Leu 7 and the  $\beta$ -hydroxy fatty acid to release the mature surfactin. Modifications blocking the reactive thiol group of the 4'-PP cofactor attached to any T domain can occur with small molecules present in the cell (acetylation, succinylation and modification with fatty acids) and pose a challenge for the organism to keep the assembly line running.

The surveillance and repair tasks for the surfactin assembly line are carried out by the stand-alone surfactin type II thioesterase

(SrfTEII)<sup>10-12</sup>. The importance of this enzyme has been demonstrated by genetic deletions that reduce the production of surfactin by 84%<sup>17</sup>. Owing to the large variety of acylation modifications and the fact that SrfTEII has to be able to interact with all seven T domains of the entire assembly line, this TEII has to be—in contrast to the TEI at the end of the last module—rather nonspecific. At the same time, premature cleavage of the correctly growing peptide chain has to be avoided. In addition to this repair function, SrfTEII might also be responsible for loading the  $\beta$ -hydroxy fatty acid onto the first C domain.

SrfTEII shows the typical  $\alpha$ -/ $\beta$ -hydrolase fold, with a central 7-stranded  $\beta$ -sheet surrounded by 8 helices (Fig. 2). Comparison with the structure of the *cis*-acting SrfTEI shows a similar overall architecture<sup>3</sup>, but with notable differences (Supplementary Fig. 5). The most obvious alteration is the amino-terminal truncation of  $\alpha$ -helix (Asp 60–Leu 67). Further modifications include the insertion of an extra helix (helix 7; Cys 193–Trp 200) between the active site residues Asp 189 and His 216, and repositioning of the helix–turn–helix motif canopying the active site in SrfTEI. This lid is moved from its central position observed in the SrfTEI crystal structure<sup>3</sup> towards the  $\beta$ -strands 6 and 7 in SrfTEII. Consequently, the active site residues in the SrfTEII structure (Ser 86, Asp 189 and His 216)<sup>10,11</sup> are just partially covered by a short loop (Gln 138–Ala 144) and are more accessible compared to SrfTEI. The space surrounding the active site is further enlarged by a kink in one of the helices of the helix–turn–helix motif.

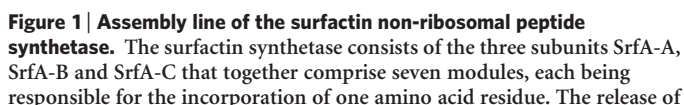
Many of the amide proton resonances of SrfTEII show line broadening or even two separate peaks, suggesting that several conformations can occur (Supplementary Fig. 2). Recently, we described the existence of several conformations in the TycC3 T domain<sup>5</sup> (the third module of the tyrocidine A synthetase subunit C, a substrate for SrfTEII) and suggested that this conformational exchange is an important driving force for the selective interaction with other domains of the NRPS system<sup>4-9,16</sup>.

NMR titration experiments<sup>18</sup> of <sup>15</sup>N-enriched SrfTEII were performed with CoA, myristoyl-CoA, apo- and holo-TycC3 T domains. The resulting chemical shift perturbations (Supplementary Fig. 3) showed interaction with the holo form and virtually no interaction with the apo form in agreement with previous experiments in which the chemical shift differences of the altered T domain forms were analysed<sup>5</sup>. Notably, myristoyl-CoA also binds to SrfTEII, suggesting recognition of this  $\beta$ -hydroxy fatty acid. This might support the previously assumed second function of SrfTEII as the starter enzyme, responsible for loading the  $\beta$ -hydroxy fatty acid onto the first C domain<sup>15</sup>. A comparison of the interaction interface of SrfTEII with

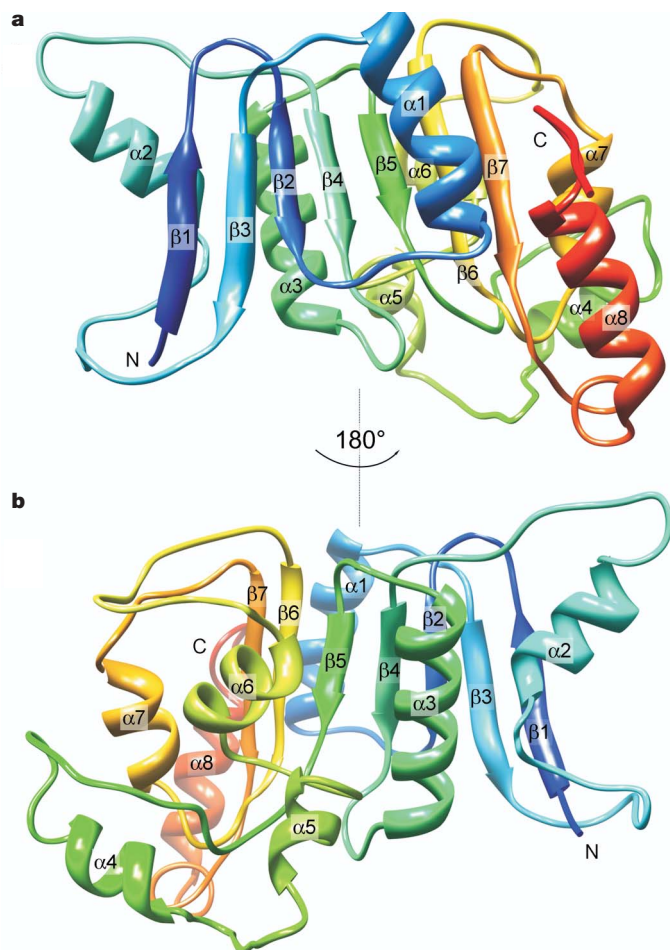
<sup>1</sup>Institute of Biophysical Chemistry and Center for Biomolecular Magnetic Resonance and Cluster of Excellence Macromolecular Complexes (CEF), J.W.-Goethe University, 60438 Frankfurt am Main, Germany. <sup>2</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>3</sup>Institute of Protein Research, 142290 Pushchino, Russia. <sup>4</sup>Department of Chemistry/Biochemistry, Philipps University, 35032 Marburg, Germany. <sup>5</sup>Frankfurt Institute for Advanced Studies, 60438 Frankfurt am Main, Germany.

To obtain more detailed insight into the interaction between SrfTEII and the TycC3 T domain, we carried out further NMR titration experiments with the acetyl-holo form of the T domain and the inactive point mutation Ser86Ala of SrfTEII (Supplementary Fig. 4). On the basis of the magnitude of the chemical shift differences, the acetyl-holo-T domain shows a stronger interaction with SrfTEII than the holo-T domain, again in agreement with previous experiments<sup>5</sup>. To determine the relative orientation of both domains, we measured nuclear Overhauser effects (NOEs) between selectively protonated methyl (Ile, Leu and Val) and aromatic (Phe) groups of SrfTEII and amide protons of the TycC3 T domain in otherwise perdeuterated proteins<sup>20,21</sup>. These NOEs and the chemical shift perturbations allowed us to calculate the structure of the complex between the TEII enzyme and the T domain (Fig. 3b). Previous investigations had revealed that the holo form of the TycC3 T domain exists in two different conformations that differ in the relative orientation of  $\alpha$ -helix 2 and the positioning of the 4'-PP cofactor<sup>5</sup>. They also demonstrated that interaction with the SrfTEII enzyme selects the more open conformation, the holo-form specific H state<sup>5</sup>. Notably, modification of the 4'-PP cofactor with an acetyl group also shifts the equilibrium and locks the T domain in the H state. Consequently, we used the H state of the TycC3 T domain in our structure calculations.

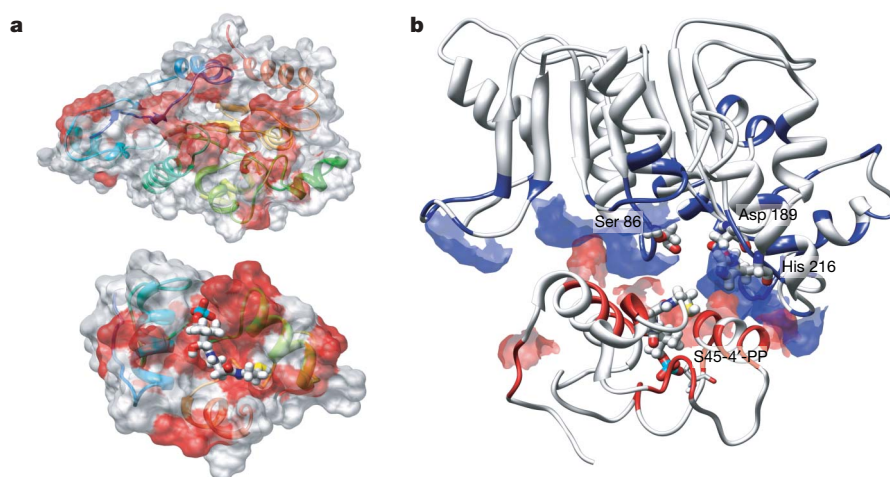
The titration experiments further demonstrated that interaction with the acetyl-holo-T domain selects one of the conformations of the thioesterase as indicated by the disappearance of the double peaks in the heteronuclear single-quantum coherence (HSQC) spectrum of the complex (Supplementary Fig. 4). Mapping the sites of the double peaks onto the structure of the TEII reveals that amino acids in the helix–turn–helix motif, loops surrounding the active site and residues located at the N termini of the first and last helix are affected (Supplementary Figs 1 and 4). Unfortunately, the regions with double peaks are separated by stretches that display only one chemical



the lipopeptide is catalysed by TEI attached to SrfC (module 7). The function of SrfD, the external thioesterase TEII (lower left), is the recycling of misprimed T domains. The 4'-PP cofactor is depicted shortened, attached to the T domains. Ac, acetyl.



**Figure 2 | Mean structure of SrfTEII.** **a, b,** Ribbon diagrams of the front (**a**) and back (**b**) views are shown. The mean structure is derived from a final ensemble of the best 20 structures out of 150 structures calculated with CYANA and refined in explicit water using the RECOORD scripts and CNS1.1. The structure is colour coded from blue (N terminus) to red (C terminus).

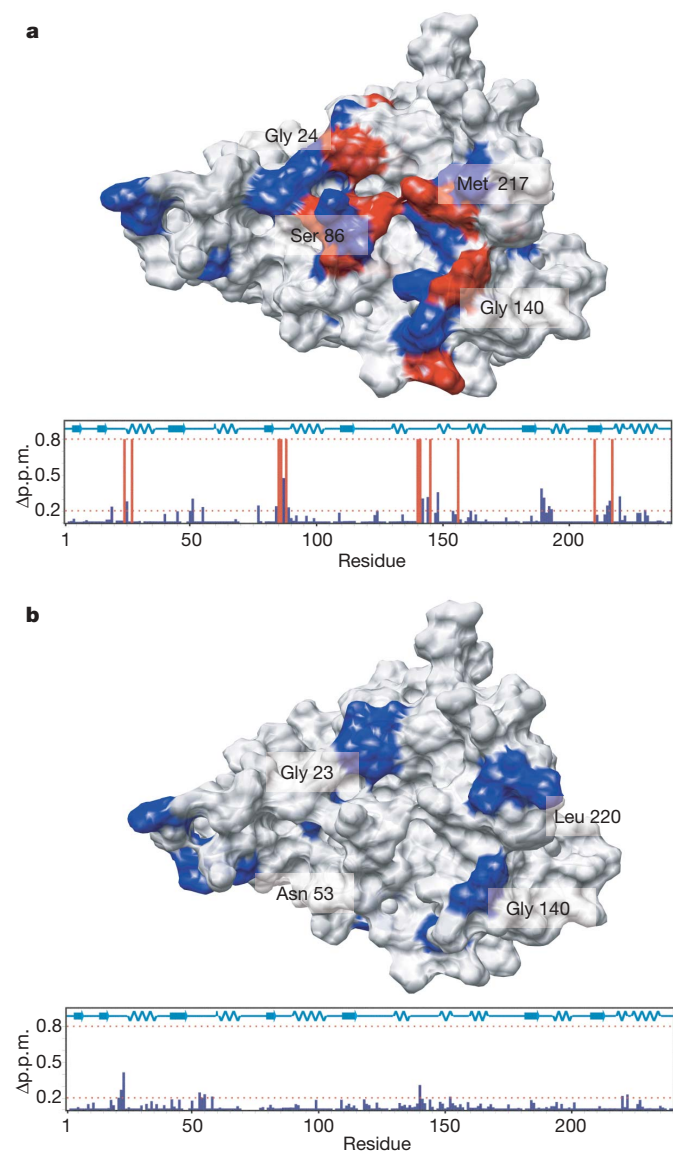


**Figure 3 | Complex structure of SrfTEII and the TycC3 T domain.** **a,** Interaction surfaces (red) for SrfTEII (top) and the TycC3 T domain (bottom) are based on chemical shift perturbations observed in NMR titration experiments of  $^{15}\text{N}$ -labelled Ser86Ala SrfTEII with unlabelled acetyl-holo-T domain and vice versa. The interaction surface identified on the T domain is identical to results published previously<sup>5</sup>. **b,** Ribbon diagram of the refined average structure of the complex of SrfTEII and the TycC3 T domain calculated using the simulated annealing protocol of CNS1.1.

shift value, making an unambiguous assignment to one particular conformation impossible. Thus, the structure of the SrfTEII represents an intermediate between two distinct conformations. The identity of the regions showing this conformational exchange combined with the selection of one state by binding of the T domain, however, allowed us to build a model of the exchange process. The two different states probably represent an open and a closed conformation of the enzyme (Supplementary Fig. 7). A similar plasticity of the lid region has been observed in the crystal structure of the SrfTEI enzyme in which two molecules with different conformations are located in the asymmetric unit<sup>3</sup>. The main difference between both molecules is the position of the lid, with one conformation representing a more open and the other a more closed state. Because both molecules, however, also form a non-native dimer in the crystal structure, the importance of this observation was not clear. The detection of an equilibrium of distinct conformations in the SrfTEII enzyme, which shifted towards one conformation by the interaction with a modified holo-T domain, is the first confirmation of such a conformational exchange process in solution. Similar exchange processes are also observed in the structure of the EntF T-TEI di-domain<sup>19</sup>.

The complex structure further shows the structural basis for the recognition of short acyl groups attached to the 4'-PP thiol. Comparison of the active sites of the TEII and TEI enzymes shows that, despite the wide opening of the active site of TEII, the space available to accommodate a group attached to the 4'-PP cofactor is limited (Fig. 3 and Supplementary Figs 6 and 8). The open conformation of the SrfTEI enzyme is characterized by a pronounced active site cavity with a volume of 630 Å<sup>3</sup>. Modelling of a heptapeptide on the basis of residual electron density has shown that this cavity is large enough to accommodate the entire peptide and to enforce a conformation that allows cyclization. In contrast to the deep, bowl-shaped cavity of the TEI enzyme, the active site of SrfTEII is embedded in a shallow groove that can accommodate only small acyl substituents on the 4'-PP cofactor (Supplementary Fig. 8). This specificity of TEIIs for small acyl substrates was previously demonstrated in kinetic studies with TycF, the TEII of the tyrocidin synthetase<sup>13</sup>, and with SrfTEII<sup>23</sup>. To investigate this selectivity further, we performed titration experiments with  $^{15}\text{N}$ -labelled SrfTEII and the unlabelled TycC3 T domain loaded with either a single amino acid (Ala) or a tripeptide (Phe-Pro-Phe). Titration with the Ala-loaded T

Surfaces are displayed only for the residues of SrfTEII (blue) and the TycC3 T domain (red) showing intermolecular NOEs. Residues showing chemical shift changes are coloured in the ribbon diagrams accordingly. Some of the active site residues of the TEII triad (Ser 86 and His 216) and the modified T domain active site (Ser 45-4'-PP) showed very severe line broadening (Supplementary Figs 3 and 4). The position of the 4'-PP cofactor shown is based on the position in the free H state of the TycC3 peptidyl carrier protein (TycC3-PCP)<sup>5</sup>.



**Figure 4 | Size limitation in substrate recognition by SrfTEII.** The combined  $^{15}\text{N}$  and  $^1\text{H}$  chemical shift perturbation ( $\Delta\text{p.p.m.}$ ) of  $^{15}\text{N}$ -HSQC-based NMR titration experiments of  $^{15}\text{N}$ -enriched SrfTEII with the unlabelled TycC3 T domain loaded either with an alanine residue non-hydrolysably bound to the 4'-PP cofactor (a) or with the tripeptide Phe-Pro-Phe (b) show a selectivity for small substrates by SrfTEII. The interaction of SrfTEII with the Phe-Pro-Phe-holo-TycC3 T domain is limited to residues also involved in holo-T domain recognition, which supports the described selection of small substrates driven by the limited size of the cavity. Residues with resonances broadened beyond detection in the 1:1 SrfTEII:Ala-holo-TycC3 T domain complex are indicated by red bars. Chemical shift differences of amino acids showing chemical shift perturbations are shown as blue bars. The surface of the SrfTEII protein is coloured accordingly.

domain showed results very similar to those with the acetyl-holo-T domain with the active site and surrounding residues showing chemical shift changes or line broadening. In contrast, the titration with the tripeptide-loaded T domain resulted in only minor chemical shift differences, far from the active site (Fig. 4). The titration results also indicate differences in the dynamic behaviour between acetyl-holo-T and Ala-holo-T on the one hand, and holo-T and peptide-loaded holo-T on the other. Whereas titrations with the first group indicate the formation of a stable complex by selection of one of the conformations represented by double peaks, titrations with the second group show a less pronounced selection of one conformation and only limited chemical shift differences.

In addition to the shape and volume of the active site, the interaction of the enzymes with the 4'-PP cofactor might also differ. In the SrfTEI structure the 4'-PP cofactor is almost completely surrounded by a channel formed by the TEI<sup>3</sup>, whereas it seems to be sandwiched between SrfTEII and the T domain in the *in trans* complex, demonstrating again the wider opening of the active site.

The structures reported here and their comparison with TEI enzymes show how modulation of the conserved thioesterase fold is used to change the function of the enzyme from one that recognizes the final product of the assembly line to one with a shallow but easily accessible active site that provides a rather unspecific but indispensable repair function.

Recently crystal structures of type I fatty acid synthetases<sup>6,24</sup> have shown the complex interaction between individual domains in these multi-enzymatic assemblies, confirming the central role that the T domains and different orientations of their cofactors have in the iterative substrate shuttling between active sites.

## METHODS SUMMARY

The C-terminal His<sub>6</sub>-tagged wild-type SrfTEII and the inactive Ser86Ala mutant proteins were heterologously expressed in *Escherichia coli* and purified by Ni-chelation affinity chromatography. All isotope enriched protein samples were produced in supplemented M9 minimal media with selectively labelled carbon and nitrogen sources. NMR spectra<sup>25–27</sup> for backbone and side chain resonance assignment, and structure calculation, were recorded on Bruker Avance800 and Avance900 spectrometers. All NMR titration experiments<sup>18</sup> of labelled protein samples with small molecules and unlabelled proteins were performed on a Bruker Avance700 spectrometer. Bruker XWINNMR or Topspin 1.3 was used for data processing and UCSF SPARKY 3.111 for resonance assignment and NOE integration. The structure of SrfTEII was calculated on the basis of 2,442 NOE upper distance limits and constraints for 92 hydrogen bonds in regular secondary structure elements using the program CYANA 2.1 (ref. 28). The 20 conformers with the lowest target function values were energy refined in explicit water using the RECOORD scripts<sup>29</sup> and the CNS 1.1 protocol<sup>30</sup>, to represent the solution structure of SrfTEII. All distance constraint violations were smaller than 0.2 Å. The *in trans* complex structure was calculated using the CNS 1.1 simulated annealing protocol. The contact surfaces of both proteins were identified by NMR titration experiments of labelled protein samples with the corresponding unlabelled interacting protein. Constraints to describe the relative domain orientation were obtained from a three-dimensional (3D)- $^{15}\text{N}$ -nuclear Overhauser enhancement spectroscopy (NOESY)-total correlation spectroscopy (TROSY) spectrum of a complex of  $^{15}\text{N}$ -labelled, perdeuterated holo-TycC3 T domain and fully perdeuterated, selectively Phe, Ile, Leu and Val protonated SrfTEII (refs 20 and 21).

**Full Methods** and any associated references are available in the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

Received 11 September 2007; accepted 10 June 2008.

- Walsh, C. T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**, 1805–1810 (2004).
- Kopp, F. & Marahiel, M. A. Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **24**, 735–749 (2007).
- Bruner, S. D. *et al.* Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**, 301–310 (2002).
- Kim, Y. & Prestegard, J. H. A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry* **28**, 8792–8797 (1989).
- Koglin, A. *et al.* Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* **312**, 273–276 (2006).
- Leibundgut, M., Jenni, S., Frick, C. & Ban, N. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* **316**, 288–290 (2007).
- Findlow, S. C., Winsor, C., Simpson, T. J., Crosby, J. & Crump, M. P. Solution structure and dynamics of oxytetracycline polyketide synthase acyl carrier protein from *Streptomyces rimosus*. *Biochemistry* **42**, 8423–8433 (2003).
- Johnson, M. A., Peti, W., Herrmann, T., Wilson, I. A. & Wüthrich, K. Solution structure of Asl1650, an acyl carrier protein from *Anabaena* sp. PCC 7120 with a variant phosphopantetheinylation-site sequence. *Protein Sci.* **15**, 1030–1041 (2006).
- Zornetzer, G. A., Fox, B. G. & Markley, J. L. Solution structures of spinach acyl carrier protein with decanoate and stearate. *Biochemistry* **45**, 5217–5227 (2006).
- Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl Acad. Sci. USA* **99**, 14083–14088 (2002).

11. Linne, U., Schwarzer, D., Schroeder, G. N. & Marahiel, M. A. Mutational analysis of a type II thioesterase associated with nonribosomal peptide synthesis. *Eur. J. Biochem.* **271**, 1536–1545 (2004).
12. Yeh, E., Kohli, R. M., Bruner, S. D. & Walsh, C. T. Type II thioesterase restores activity of a NRPS module stalled with an aminoacyl-S-enzyme that cannot be elongated. *ChemBioChem* **5**, 1290–1293 (2004).
13. Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174–4183 (1997).
14. Gehring, A. M., Mori, I. & Walsh, C. T. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37**, 2648–2659 (1998).
15. Steller, S. *et al.* Initiation of surfactin biosynthesis and the role of the SrfD-thioesterase protein. *Biochemistry* **43**, 11331–11343 (2004).
16. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L.-O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781–792 (2007).
17. Schneider, A. & Marahiel, M. A. Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch. Microbiol.* **169**, 404–410 (1998).
18. Zuiderweg, E. R. P. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* **41**, 1–7 (2002).
19. Frueh, D. P. *et al.* Dynamic thiolation–thioesterase structure of a non-ribosomal peptide synthetase. *Nature* doi:10.1038/nature07162 (this issue).
20. Gardner, K. H. & Kay, L. E. Production and incorporation of  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^2\text{H}$  ( $^1\text{H}$ - $\delta^1$  methyl) isoleucine into proteins for multidimensional NMR studies. *J. Am. Chem. Soc.* **119**, 7599–7600 (1997).
21. Frueh, D. P., Vosburg, D. A., Walsh, C. T. & Wagner, G. Determination of all NOEs in  $^1\text{H}$ - $^{13}\text{C}$ -Me-ILV-U- $^2\text{H}$ - $^{15}\text{N}$  proteins with two time-shared experiments. *J. Biomol. NMR* **34**, 31–40 (2006).
22. Lai, J. R., Koglin, A. & Walsh, C. T. Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry* **45**, 14869–14879 (2006).
23. Schwarzer, D., Mootz, H. D., Linne, U. & Marahiel, M. A. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proc. Natl Acad. Sci. USA* **99**, 14083–14088 (2002).
24. Jenni, S. *et al.* Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254–261 (2007).
25. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated  $T_2$  relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl Acad. Sci. USA* **94**, 12366–12371 (1997).
26. Sattler, M., Schleucher, J. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).
27. Ferentz, A. E. & Wagner, G. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q. Rev. Biophys.* **33**, 29–65 (2000).
28. Güntert, P., Mumenthaler, C. & Wüthrich, K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298 (1997).
29. Nederveen, A. J. *et al.* RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. *Proteins* **59**, 662–672 (2005).
30. Brünger, A. T. *et al.* Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).

**Supplementary Information** is linked to the online version of the paper at [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank M. Strieker for editing the manuscript and B. Schaefer for her help and support in sample preparation. We thank Chi Scientific Inc. for the fast and high quality supply of the substrate peptides. The research was funded by the research grant BE-19/11 of the Deutsche Forschungsgemeinschaft (F.B. and M.A.M.); an enclosed fellowship (A.K.); the Centre for Biomolecular Magnetic Resonance at the University Frankfurt; the Cluster of Excellence Frankfurt (Macromolecular Complexes); and the Volkswagen Foundation (P.G.). A.K. thanks the Human Frontier Science Program Organization for a long-term fellowship awarded in April 2007.

**Author Contributions** A.K., V.D., F.B. and M.A.M. designed the experiments and defined the research for the SrfTEII. A.K., V.D. and C.T.W. wrote the manuscript and designed the experiments for the complex structure. A.K., F.B. and F.L. conducted the research including protein expression, data acquisition, resonance assignment and structure calculation. V.R.R. helped with resonance assignment and structure calculation of the SrfTEII. P.G. provided the volumes of the cavities of SrfTEI and of the complex SrfTEII–TycC3 T domain and supported the complex structure calculation. D.P.F. and G.W. recorded the NMR spectra of the protein complex. A.K. calculated the structure of the isolated SrfTEII, the structure of the protein complex (SrfTEII and TycC3–PCP) and performed all titration experiments. M.A.M. and M.R.M. provided the expression vectors for SrfTEII and the TycC3 T domain. M.A.M. provided the purification protocol and biochemical characterization of the SrfTEII. E.R.S. provided the non-hydrolysable CoA derivative and synthesized the peptide and amino acid modified substrate analogues.

**Author Information** The coordinates have been deposited in the Protein Data Bank under accession number 2RON (structure of SrfTEII) and 2K2Q (complex structure of SrfTEII and H state TycC3–PCP). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). Correspondence and requests for materials should be addressed to V.D. (VDoetsch@em.uni-frankfurt.de), C.T.W. (Christopher\_Walsh@hms.harvard.edu) or M.A.M. (Marahiel@chemie.uni-marburg.de).

## METHODS

**Protein expression and purification.** The SrfTEII protein (residues 1–241) and its inactive Ser86Ala mutant were heterologously expressed in the *E. coli* strain BL21 (DE3) using a pET-expression vector system containing a C-terminal His<sub>6</sub>-tag. They were purified by Ni-chelation affinity chromatography and subsequent gel filtration chromatography using a Pharmacia Superdex-75 column. The purity of all protein samples was validated by SDS-PAGE analysis. All labelled samples of the SrfTEII protein were produced in supplemented M9 media with stable isotope enriched glucose (<sup>13</sup>C or <sup>13</sup>C/<sup>2</sup>H; Cambridge Isotopes Laboratories) as the only carbon source and <sup>15</sup>N ammonium chloride (Cambridge Isotopes Laboratories) as the nitrogen source. For the preparation of fully perdeuterated samples the aqueous solvent was replaced by D<sub>2</sub>O and perdeuterated glucose was used as the carbon source. The stable isotope enriched (<sup>2</sup>H, <sup>15</sup>N, <sup>13</sup>C) and unlabelled samples of the TycC3 T domain (1–87) were expressed and purified as described previously<sup>5</sup>.

**NMR data acquisition and assignment.** All NMR spectra for backbone and side chain resonance assignment and structure calculation of the SrfTEII protein were recorded on an Avance800 or an Avance900 spectrometer equipped with a 5 mm triple resonance, z-gradient cryogenic probe at 298 K. The resonance assignment was based on standard triple-resonance experiments, following standard protocols<sup>25–27,31–33</sup>. An additional 4D <sup>15</sup>N-<sup>15</sup>N-resolved NOESY spectrum of <sup>15</sup>N-labelled and 70% perdeuterated SrfTEII was recorded to verify the sequential resonance assignment and to define unambiguous amide–amide based distance constraints. DSS (4, 4-dimethyl-4-silapentane-1-sulphonate) was used as an internal chemical shift reference. All <sup>15</sup>N-HSQC based NMR titration experiments<sup>18</sup> were performed on an Avance700 spectrometer at 296 K. The spectral width was set to 12.5 p.p.m. in the proton dimension and 35 p.p.m. in the nitrogen dimension. A total of 2,048 points in the direct and 1,024 points in the indirect dimension were collected for all HSQC spectra. The TROSY-based <sup>15</sup>N-edited NOESY experiments to measure interface distance constraints in the complex of Ser86Ala SrfTEII with the acetyl-holo-TycC3 T domain (1:1 ratio) and wild type SrfTEII with the holo-TycC3 T domain (1:2.5 ratio) were recorded on an Avance900 with a mixing time of 182 ms. XWINNMR 3.1 or Topspin 1.3 (Bruker) were used for processing and SPARKY 3.111 (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California) for resonance assignment and NOE peak integration. The SrfTEII protein could be assigned to 92% completeness of all 227 non-proline backbone resonances and 91.5% of all resonances.

**Structure calculation and refinement.** Structures of SrfTEII were calculated by simulated annealing in torsion angle space with the program CYANA 2.1 (refs 28 and 34). Refinements in explicit water (TIP3P) were performed using the RECOORD scripts<sup>29</sup> and the CNS 1.1 protocol<sup>30</sup>. In all calculations 210 unambiguously identified backbone amide to amide contacts from a 4-dimensional <sup>15</sup>N/<sup>15</sup>N-resolved NOESY experiment<sup>26</sup> were applied with upper distance bounds of 6.5 Å. The secondary structure was defined by 184 distance constraints for 92 backbone hydrogen bonds identified on the basis of unambiguous NOEs in a 3D-<sup>15</sup>N-resolved NOESY spectrum and by 278 torsion angle constraints derived from chemical shift values with the program TALOS<sup>35</sup>. Furthermore, 2,442 distance constraints were generated on the basis of 7,802 NOESY cross peaks from the same 3D-<sup>15</sup>N-resolved NOESY and three 3D-<sup>13</sup>C-resolved NOESY spectra recorded for aliphatic, methyl and aromatic side chains separately. An initial structure was calculated from 2,851 manually assigned NOEs and the torsion angle constraints with CYANA 2.1. The calculation yielded a structural bundle (20 out of 100 calculated structures, sorted by lowest target function value) with a precision (root mean square deviation, r.m.s.d.) of 3.8 Å and an average target function value of 1.9 Å<sup>2</sup>. The CYANA script with automatic NOE assignment was used with the resulting 1,706 distance constraints of the previous calculation and an extra 94 distance constraints obtained from manually assigned CH-NOEs selective for aromatic side chains. The interpretation of 4,951 ambiguous NOEs resulted in an additional 642 distance constraints and in a structural bundle with an r.m.s.d. of 1.35 Å for all heavy atoms and an average target function value of 4.9 Å<sup>2</sup> for the 20 structures with lowest CYANA target function values out of 150 calculated conformers. The final set of 20 out of 150 calculated structures does not show distance or van der Waals violations larger than 0.20 Å, no angle violations >2.6° and 75% of all dihedral angles are located in most favoured regions and 24.8% in additionally allowed regions of the Ramachandran plot. Pockets in SrfTEI (ref. 3) and SrfTEII were identified and their sizes computed using the CASTp algorithm<sup>36,37</sup> with a probe radius of 2.0 Å.

**NMR titrations and complex structure calculation.** Chemical shift perturbations were measured in <sup>15</sup>N-HSQC-based NMR spectra for titration experiments<sup>18</sup> of <sup>15</sup>N-labelled holo- and acetyl-holo-TycC3 T domain with unlabelled SrfTEII and vice versa. The chemical shift perturbations and the line shapes of all titration experiments were analysed using MestReC 4.9.9.6 (ref. 38).

It has been demonstrated previously that the H state of the TycC3 T domain is recognized as a substrate by the SrfTEII (ref. 5). Interaction surfaces of the carrier protein and the thioesterase were identified by the titration experiments and the structures of both proteins were used to calculate the structure of the enzymatically active complex. Constraints to describe the relative domain orientation were obtained from a 3D-<sup>15</sup>N-resolved NOESY-TROSY spectrum of a <sup>15</sup>N-labelled completely perdeuterated the holo-TycC3 T domain and fully perdeuterated selectively Phe, Ile, Leu and Val protonated SrfTEII (refs 20 and 21). Nine unambiguous and seventeen additional ambiguous constraints between the TycC3 T domain amide protons and SrfTEII FILV-side-chain protons were identified. Distance constraints on the basis of these NOEs were applied to the structural calculation of the *in trans* di-domain complex using the CNS 1.1 protocol. All structural figures were prepared using UCSF CHIMERA 1.2470 (ref. 39).

**In vitro modification of the TycC3 T domain.** The *in vitro* modification of the unlabelled apo-TycC3 T domain was carried out in 2 ml reaction mixtures of 0.25 mM apo-TycC3 T domain, 0.5 mM acetyl-CoA (Sigma), 20 μM Sfp and 5 mM MgCl<sub>2</sub>, buffered in 100 mM sodium phosphate at pH 8.0 for 45 min at room temperature. The reaction mixture was subsequently purified by desalting (Econo-pac 10 DG desalting column (BioRad)) and concentrated using an Amicon Ultra 4 Ultracell – 5k (Millipore) filter device with a molecular weight cutoff of 5 kDa.

**Synthesis of peptidyl-amino CoA substrates.** The general procedure for the synthesis of peptidyl-amino CoA substrates was based on a synthesis described previously<sup>40,41</sup>. In brief, 10 mmol of amino-CoA<sup>42</sup>, 15 mmol of PyBOP and 40 mmol of potassium carbonate were added to 10 mmol of the Boc-protected peptide. The solids were subsequently dissolved in a 1:1 THF/water mixture (500 ml total) and allowed to stir at room temperature overnight. The reaction mixture was directly purified by preparative high-performance liquid chromatography (HPLC) using a single injection on a Phenomenex C18 250 × 21.2 mm, 10 mm, 100 Å column and eluting using a gradient of 0% to 60% acetonitrile containing 0.1% trifluoroacetic acid over 30 min and a flow rate of 10 ml min<sup>−1</sup> while monitoring at 260 nm. The identities of the Boc-protected peptidyl-amino CoA substrates were verified by HPLC-mass spectrometry (MS) and matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry. Next, cleavage of the Boc-protecting group was carried out by dissolving the Boc-protected peptidyl-amino CoA in a 95:2.5:2.5 mixture of trifluoroacetic acid, trifluoroethanol and water, and allowing this mixture to stir at room temperature for 2 h. The deprotected peptidyl-amino CoA substrates were purified by preparative HPLC using the same conditions as described above. HPLC-MS and MALDI-TOF MS were used to confirm the identity of the peptidyl-amino CoA substrates.

- Ikura, M., Kay, L. E. & Bax, A. Improved three-dimensional <sup>1</sup>H-<sup>13</sup>C-<sup>1</sup>H correlation spectroscopy of a <sup>13</sup>C-labeled protein using constant-time evolution. *J. Biomol. NMR* **1**, 299–304 (1991).
- Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl Acad. Sci. USA* **95**, 13585–13590 (1998).
- Bax, A., Clore, M. & Gronenborn, A. M. <sup>1</sup>H-<sup>13</sup>C Correlation via isotropic mixing of <sup>13</sup>C magnetization, a new three-dimensional approach for assigning <sup>1</sup>H and <sup>13</sup>C spectra of <sup>13</sup>C-enriched proteins. *J. Magn. Reson.* **88**, 425–431 (1990).
- Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278**, 353–378 (2004).
- Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
- Liang, J., Edelsbrunner, H. & Woodward, C. Anatomy of protein pockets and cavities: Measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**, 1884–1897 (1998).
- Dundas, J. et al. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, W116–W118 (2006).
- Cobas, C. F. & Sardina, J. Nuclear magnetic resonance data processing. MestRe-C: A software package for desktop computers. *Concepts Magn. Reson.* **A 19**, 80–96 (2003).
- Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489 (1999).
- Sieber, S. A., Walsh, C. T. & Marahiel, M. A. Loading peptidyl-coenzyme A onto peptidyl carrier proteins: a novel approach in characterizing macrocyclization by thioesterase domains. *J. Am. Chem. Soc.* **125**, 10862–10866 (2003).
- Meier, J. L., Mercer, A. C., Rivera, H. & Burkart, M. D. Synthesis and evaluation of bioorthogonal pantetheine analogues for *in vivo* protein modification. *J. Am. Chem. Soc.* **128**, 12174–12184 (2006).

# Subject to change

Getting to grips with scientific variables.

## Joseph Lachance

Huang, L. *et al.* Comments re: Lee, V. *et al.* *Science* **322**, 123 (2009).

"All efforts to repeat the results by Lee *et al.* have been unsuccessful. While it is understandable that they lack our impeccable experimental technique, we believe that they failed to properly control for..."

Jones, M. C. Irreproducibility. *Philos. Sci.* **78**, 143-151 (2011).

"An increasing number of studies show a lack of reproducibility. In addition, accusations of scientific fraud have skyrocketed in the past year. Most famously, the dry-ice bomb incident between chemists at Oxford and Chicago led to no less than 7 arrests..."

Jimenez, F. Physical laws are local. *Nature* **482**, 432-435 (2014).

"Research sites on either side of the Wallace Line consistently obtain different results. These experiments conclusively demonstrate that physical laws are not the same in different locations. It has not escaped our attention that this phenomenon may be responsible for numerous discrepancies in the past 5 years<sup>9-17</sup>. We have no choice but to conclude: what happens in Vegas, stays in Vegas..."

Lapointe, A. The End of science? *Time* **186**, 24-31 (2015).

"This is not your father's or mother's physics. At a recent conference in Zurich, physicists heatedly debated the objectivity of science. If physical laws do indeed vary from location to location, the foundations of science are on shaky ground. Matters were not helped when it was revealed that a cultural relativist accepted the Ultimate Boeing 747 gambit, only to glide safely and slowly down to earth from a height of 30,000 feet. Science as defined by Karl Popper is dead..."

Orbach, R. *et al.* Aether or dark matter? *Astrophys. J.* **814**, 78-90 (2016).

"Signals from the pulsar J1342+2822B seem to have been modified in transit. Results are consistent with the effect of a gravitational lens, perhaps due to the existence of dark matter. However, an alternative explanation is that the signal passed through a region of space with different physical laws. If so, this would mean that

the Jimenez effect is more pervasive than previously thought..."

Kabaka, F. *et al.* Evidence of dynamic physical laws. *PLoS Phys.* **6**, 1-17 (2017).

"Results from supercollider experiments exhibited a marked change after 09:43 GMT on 12 January 2017. Something clearly changed. Prior to 12 January, readings at CERN were  $9.34 \pm 0.3$  MeV. Subsequent readings were  $5.34 \pm 0.2$  MeV. Interestingly, readings at Fermilab were consistently in the 9.34 MeV range. This suggests that physical laws are both local and dynamic..."



Singh, N. A history of cellular automata. *Phys. Rev. Lett.* **118**, 56-72 (2017).

"Furore over dynamic physical laws has led to increased interest in cellular automata and artificial life. Cellular automata are spatial models in which states of cells within a grid are updated in discrete steps. Simple spatial rules, such as those of Conway's Game of Life, can generate complex results. Here, the state of a cell depends on the states of neighbouring cells. Patterns such as 'blocks', 'blinkers' and 'gliders' have been found. Some patterns give the appearance of being alive. It is rumoured that the amount of computing time dedicated to the Game of Life is exceeded only by SETI@Home and World of Warcraft..."

Chen, X. & Li, H. Meta-rules and the search for universality. Preprint at <<http://arxiv.org/abs/0992.4174>> (2019).

"Recent discoveries have divided the scientific community into two warring camps. Localists believe the new paradigm that physical laws are local, whereas Universalists cling to the idea that physical laws are

universal. We suggest that local laws are determined by some universal meta-rule. This could bring the Localist-Universalist feud to an end, allowing resources to be diverted to empirical studies of Murphy's law (a research programme that has garnered such prestigious awards as the Ig Nobel prize)..."

Williams, C. *et al.* Physical Laws are Cellular Automata. *Proc. Natl Acad. Sci. USA* **116**, 78-92 (2019).

"As shown in Figure 3, local laws are spatially and temporally correlated. Laws belong to a finite number of configurations and depend on laws present at nearby locations. This indicates that physical laws behave as cellular automata. Patterns of physical laws exhibit both periodicity and mobility. Interestingly, 'Rule 110' seems to govern the behaviour of fringe elements..."

Abel, C. & Wilkins, O. Evolution of simulated organisms. *Artif. Life* **26**, 231-239 (2020).

"Artificial life can exist in the form of cellular automata. Using a set of iterated, spatial rules, we simulated an ecosystem on a large cluster of microcomputers. For the first time, the existence of a virtual organism more complex than *Escherichia coli* has been demonstrated (see supplemental pages S1-S23 for definitions of complexity). It remains to be seen if sentience can evolve *in silico*..."

Meme, A. Life within the laws of the Universe. [www.stochasticity.scienceblogs.com](http://www.stochasticity.scienceblogs.com) (2021).

"If I understand things correctly, this means that life, or something like it, exists on a truly grand scale. No, the Universe is not a giant game of Go. Instead, the rules of the Universe are flickering between states as if they are cellular automata. The bizarre events of the past few years seem to fit with a 'glider' passing through our local neighbourhood. Things have returned to normal, but for how long? If 'gliders' can exist, so too might other more complex patterns. Does this mean that the Universe itself is alive? I can only begin to wonder what strange organisms exist within the fabric of the Universe. Perhaps we are not alone!..." ■

Joseph Lachance is a PhD student at Stony Brook University, where he studies theoretical population genetics and *Drosophila* genetics.

JACEY